



M 2014

SELEÇÃO DE VARIÁVEIS EM APLICAÇÕES DE PREVISÃO

JOSÉ MIGUEL SANTOS SOUSA

DISSERTAÇÃO DE MESTRADO APRESENTADA

À FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO EM 30/07/2014

ENERGIA

Faculdade de Engenharia da Universidade do Porto



Seleção de Variáveis em Aplicações de Previsão

José Miguel Santos Sousa

Dissertação realizada no âmbito do
Mestrado Integrado em Engenharia Eletrotécnica e de Computadores
Major Energia

Orientador: Prof. José Nuno Moura Marques Fidalgo

Julho de 2014

A Dissertação intitulada

“Seleção de Variáveis em Aplicações de Previsão”

foi aprovada em provas realizadas em 16-07-2014

o júri



Presidente Professora Doutora Maria Helena Osório Pestana de Vasconcelos
Professora Auxiliar do Departamento de Engenharia Eletrotécnica e de
Computadores da Faculdade de Engenharia da Universidade do Porto



Professora Doutora Teresa Alexandra Ferreira Mourão Pinto Nogueira
Professor Adjunto do Departamento de Engenharia Electrotécnica do Instituto
Superior de Engenharia do Porto



Professor Doutor José Nuno Moura Marques Fidalgo
Professor Associado do Departamento de Engenharia Eletrotécnica e de
Computadores da Faculdade de Engenharia da Universidade do Porto

O autor declara que a presente dissertação (ou relatório de projeto) é da sua exclusiva autoria e foi escrita sem qualquer apoio externo não explicitamente autorizado. Os resultados, ideias, parágrafos, ou outros extratos tomados de ou inspirados em trabalhos de outros autores, e demais referências bibliográficas usadas, são corretamente citados.



Autor - José Miguel Santos Sousa

Faculdade de Engenharia da Universidade do Porto

Resumo

A seleção de variáveis é um dos procedimentos mais comuns em diversas aplicações de previsão ou outras tarefas que envolvem processos de regressão. Em muitos casos, as variáveis de entradas são escolhidas com base em critérios subjetivos, tais como a sensibilidade do utilizador, ou outros critérios discricionários.

Neste trabalho procura-se encontrar uma alternativa objetiva e não-discricionária ao problema de seleção de variáveis, ou seja, que não esteja condicionada pelo fator humano, mas sim apoiada por conceitos matemáticos. É realizada uma revisão de alguns conceitos teóricos relacionados com a Entropia, incluindo o conceito de Informação Mútua, sendo depois proposta uma solução ao nosso problema de seleção de variáveis.

O caso de estudo consistiu num problema de previsão de Consumo Ibérico, com base em dados reais registados ao longo de um período de aproximadamente 3 anos. A ferramenta de regressão foi baseada em redes neuronais, dado o seu sucesso neste domínio. Este caso de teste pretende determinar as qualidades e fragilidades da metodologia proposta para seleção de variáveis. Os resultados apresentados visam sobretudo ilustrar o potencial desta abordagem, através da análise da coerência dos desempenhos de previsão obtidos com diferentes combinações de variáveis de entrada. A ideia principal consistiu em obter uma ligação entre os índices obtidos pela aplicação do conceito de Informação Mútua e as performances de previsão obtidas, para diferentes combinações de entradas.

Palavras-Chave:

- Entropia; Teoria da Informação; Informação Mútua; Redes Neuronais.

Abstract

In this paper we try to find a reliable alternative to the variable selection problem, one that isn't affected by the human factor, but supported by proven mathematical concepts. We start by reviewing some theoretical concepts, such as Entropy, Neural Networks and Mutual Information, and then proceeding with a solution proposal for this problem that is variable selection. This solution will be applied to a Load Forecasting problem of the Iberian Peninsula, through the use of approximately 3 years' worth of data. This test case will allow us to draw any conclusions of the applied method, and its utility in a variable selection scenario. The results will then be tested, through the use of Neural Networks, where we try to find a link between the Mutual Information variable's values, and the Neural Networks performance, using those same variables.

Key-Words:

- *Entropy; Information Theory; Mutual Information; Neural Networks.*

Agradecimentos

Quero deixar um agradecimento especial ao Professor José Nuno Marques Fidalgo, por todo o seu apoio, sugestões e correções ao longo de todo este trabalho. Por ser capaz de fornecer uma visão diferente dos resultados quando estes não eram tão bons quanto se esperava, e das suas potencialidades quando o eram.

Agradecer aos meus pais, por todo o apoio que me deram durante todos estes anos, especialmente nos maus momentos.

Um agradecimento especial ao Professor Cláudio Domingos Martins Monteiro e à empresa SmartWatt pela autorização de utilização dos dados que permitiram o desenvolvimento deste trabalho.

Queria agradecer a todos aqueles que contribuíram, direta ou indiretamente, para o desenvolvimento deste trabalho.

Índice

Resumo	iv
Abstract.....	vii
Agradecimentos	ix
Índice.....	xi
Lista de figuras	xiii
Lista de tabelas	xiv
Lista de equações.....	xvi
Abreviaturas e Símbolos	xvii
Capítulo 1	1
Introdução.....	1
1.1 - Motivação e Objectivos	2
1.2 - Contribuições/Ferramentas	2
1.3 - Estrutura da Dissertação.....	3
Capítulo 2	4
Metodologia	4
Capítulo 3	5
Conceitos e Ferramentas	5
3.1 - Entropia	5
3.1.1 - Entropia e a teoria da informação	6
3.1.2 - Propriedades da Entropia	7
3.1.3 - Entropia conjunta e entropia condicional	7
3.1.4 - Informação Mútua	8
3.2 - Redes Neurais	9
3.2.1 - Aprendizagem de uma rede neuronal	11
3.2.1.1 - Aprendizagem por Retropropagação do erro	11
3.2.2 - Vantagens e Desvantagens da utilização de redes neurais	12
Capítulo 4	13
Aplicação do Conceito de Informação Mútua a um caso prático	13
4.1 - Tratamento da Informação.....	13

4.1.1 - Construção da matriz das probabilidades	15
4.2 - Cálculo dos índices de Informação Mútua	16
4.3 - Evolução do índice de Informação Mútua	17
4.4 - Manipulação das variáveis discretas	21
4.4.1 - Variável Hora_168	21
4.4.2 - Variável DS_8h	22
4.5 - Variáveis Meteorológicas	23
Capítulo 5	25
Testes de Desempenho	25
5.1 - Criação e análise de redes com uma variável de entrada	25
5.2 - Criação e análise de redes com múltiplas entradas	27
5.3 - DAConsumo vs SAConsumo	29
5.4 - Teste das variáveis Manipuladas	32
5.4.1 - Teste da variável Hora_168	32
5.5 - Teste das variáveis Meteorológicas	36
Capítulo 6	38
Conclusão	38
Referências	40
Anexos	43
Anexo A	43
Anexo B	45
Anexo C	46

Lista de figuras

Figura 1 - Entropia esperada do lançamento de uma moeda justa.....	6
Figura 2 - Diagrama de Venn da Informação Mútua	9
Figura 3 - Modelo não linear de um neurónio	10
Figura 4 - Representação geral da estrutura de uma rede neuronal.....	10
Figura 5 - Representação gráfica da evolução dos valores da variável DAConsumo com a dimensão da matriz das probabilidades.....	18
Figura 6 - Representação gráfica da evolução dos valores da variável SAConsumo com a dimensão da matriz das probabilidades.....	18
Figura 7 - Representação gráfica da evolução dos valores da variável Temperatura com a dimensão da matriz das probabilidades.....	19
Figura 8 - Representação gráfica da evolução dos valores da variável Pluviosidade com a dimensão da matriz das probabilidades.....	19
Figura 9 - Representação gráfica da evolução dos valores da variável Vel. Vento com a dimensão da matriz das probabilidades.....	19
Figura 10 - Representação gráfica da evolução dos valores da variável Irradiância com a dimensão da matriz das probabilidades.....	20
Figura 11 - Representação gráfica do MSE em função da razão IM, para redes neuronais com 1 entrada	27
Figura 12 - Representação gráfica do MSE em função da razão IM, para redes neuronais c/ DS e Hora no conjunto de entradas	28

Lista de tabelas

Tabela 1- Matriz Probabilidades DS vs Consumo (em %)	15
Tabela 2 - Coeficientes de IM para dimensões 8 e 24	16
Tabela 3 - Importância relativa das variáveis de acordo com os índices de IM para dimensões 8 e 24.....	16
Tabela 4 - Evolução dos índices de IM de acordo com a dimensão da matriz das probabilidades	17
Tabela 5 - Evolução da importância relativa das variáveis de acordo com os índices de IM.....	17
Tabela 6 - Evolução dos valores da variável DAConsumo com a dimensão da matriz das probabilidades	18
Tabela 7 - Evolução dos valores da variável SAConsumo com a dimensão da matriz das probabilidades	18
Tabela 8 - Evolução dos valores da variável Temperatura com a dimensão da matriz das probabilidades	19
Tabela 9 - Evolução dos valores da variável Pluviosidade com a dimensão da matriz das probabilidades	19
Tabela 10 - Evolução dos valores da variável Vel. Vento com a dimensão da matriz das probabilidades	18
Tabela 11 - Evolução dos valores da variável Irradiância com a dimensão da matriz das probabilidades	20
Tabela 12 - Coeficientes de IM para a dimensão 168.....	22
Tabela 13 - Importância relativa das variáveis de acordo com os índices de IM para a dimensão 168	22
Tabela 14 - Coeficientes de IM para a dimensão 24 c/ a variável DS_8h	23
Tabela 15 - Importância relativa das variáveis de acordo com os índices de IM pra a dimensão 24 c/ a variável DS_8h.....	23

Tabela 16 - Índices de IM das variáveis meteorológicas	24
Tabela 17 - Índices de IM das variáveis meteorológicas medidas 2 dias antes.....	24
Tabela 18 - Valores do MAPE e MSE das redes neuronais c/ 1 entrada.....	26
Tabela 19 -Comparação entre os índices de IM, e os erros obtidos pelas redes neuronais	26
Tabela 20 -Comparação entre os índices de IM e o MSE c/ DS e Hora no conjunto de entradas	18
Tabela 21 - Comparação entre os índices IM e o MSE e MAPE: Hora, DS, SAConsumo vs Hora, DS, DAConsumo	30
Tabela 22 - Comparação entre os índices de IM e o MSE: SAConsumo vs DAConsumo	30
Tabela 23 - Comparação entre os índices de IM e o MSE e MAPE: SAConsumo vs DAConsumo c/ Temperatura	31
Tabela 24 - Valores do MAPE e MSE das variáveis Hora, DS e Hora_168	32
Tabela 25 - Valores do MAPE e MSE p/ conjuntos de entrada c/ Hora e DS	33
Tabela 26 - Valores do MAPE e MSE p/ conjuntos de entrada c/ Hora_168	33
Tabela 27 - Valores do MAPE e MSE das variáveis DS e DS_8h	34
Tabela 28 - Valores do MAPE e MSE das variáveis DS+Hora e DS_8h	34
Tabela 29 - Valores do MAPE e MSE c/ DS	35
Tabela 30 - Valores do MAPE e MSE c/ DS_8h.....	35
Tabela 31 - Valores do MAPE e MSE c/ DS e Hora	35
Tabela 32 - Valores do MAPE e MSE c/ DS_8h e Hora.....	35
Tabela 33 - Valores do MAPE e MSE para comparação de DS e DS_8h	35
Tabela 34 - Valores do MAPE e MSE das variáveis meteorológicas	36
Tabela 35 - Valores do MAPE e MSE das variáveis meteorológicas 2 dias antes	36
Tabela 36 - Valores do MAPE e MSE das variáveis meteorológicas c/ Hora	36
Tabela 37 - Valores do MAPE e MSE das variáveis meteorológicas 2 dias antes c/ Hora	36

Lista de equações

Equação 3.1 - Entropia de Shannon.....	6
Equação 3.2 - Entropia de uma moeda não viciada.....	6
Equação 3.3 - Expressão generalizada da Entropia de Shannon, para 2 variáveis.....	6
Equação 3.4 - Entropia Condicional de Shannon	6
Equação 3.5 - Entropia Condicional de Shannon, generalizada para todo o conjunto X.....	6
Equação 3.6 - Propriedades da entropia	6
Equação 3.7 - Propriedade genérica da entropia.....	6
Equação 3.8 - Informação Mútua (variáveis discretas)	6
Equação 3.9 - Informação Mútua (variáveis contínuas)	6

Abreviaturas e Símbolos

Lista de abreviaturas (ordenadas por ordem alfabética)

RN	Rede Neuronal
IM	Informação Mútua
MP	Matriz das Probabilidades
Nftool	<i>Neural network fitting tool</i>
FEUP	Faculdade de Engenharia da Universidade do Porto

Lista de símbolos

H	Entropia
I	Quantidade de informação
p_i	Probabilidade do acontecimento i
$r(x_i, y_i)$	Probabilidade conjunta dos acontecimentos x_i e y_i

Capítulo 1

Introdução

O problema da seleção de variáveis pode não parecer algo significativo, nada que o processo de tentativa-erro não possa solucionar, mas quando se analisa a sua importância em previsões complexas, com quantidades elevadas, tanto de informação, como de variáveis nos conjuntos de entrada e saída, em que o tempo de computação é algo a se ter em conta, o caso muda de figura.

Não se pode dar ao luxo de, em certas previsões, se usar as variáveis erradas. Previsões a curto-prazo e/ou de grande importância têm que usar, sem sombra de dúvida, as variáveis que irão proporcionar os melhores resultados para o sistema em questão.

No entanto, esta tarefa nem sempre é fácil, sendo fácil resignarmo-nos às experiências passadas, em vez de efetuar uma análise mais detalhada da situação.

Algumas ferramentas existentes, e mais comuns, também têm as suas limitações. A correlação por exemplo, apenas pode ser usada em sistemas onde a relação entre a entrada e saída é linear. Se tal não acontecer a indicação dada é errada, e pode induzir o previsor em erro.

Surge então a necessidade de encontrar novos métodos, capazes de distinguir as melhores variáveis para cada sistema, não através de medidas subjetivas (fator humano), mas através de conceitos matemáticos provados e sustentados.

Neste trabalho analisaremos um método encontrado, o conceito da Informação Mútua, ao qual o aplicaremos a um problema de previsão de carga, algo que já foi muito estudado, e do qual já se tem bastante conhecimento. Este é no entanto o caso de teste, pois o interesse é que o método seja possível de ser implementado a qualquer sistema de previsão.

Após a revisão de alguns conceitos teóricos relacionados com os conceitos de Entropia e Informação Mútua, iremos aplicar estes conceitos ao nosso caso de estudo. Os resultados obtidos serão analisados tendo em conta as variáveis utilizadas, sendo depois realizados alguns testes, de forma a comprovar ou desmentir os resultados obtidos.

1.1 - Motivação e Objetivos

A previsão de acontecimentos encontra-se em toda a nossa volta. Desde a previsão do estado do tempo para amanhã, até à previsão de preços de variadíssimos produtos, passando é claro pela previsão de cargas, não existe limite para o que se pretende prever.

Estas previsões estão no entanto muitas vezes sujeitas a erros grosseiros, devido a uma incorreta utilização de variáveis, o que leva a aumento de erros de previsão, e desperdícios de tempo na computação.

Existe então todo o interesse em encontrar métodos que nos auxiliem neste processo, e que nos deem a indicação de quais as melhores variáveis a utilizar num certo caso, e quais poderão ser descartadas, permitindo-nos chegar a resultados mais próximos da realidade, num menor espaço de tempo, através da diminuição, não só do número de variáveis a considerar, como do número de combinações das mesmas.

O objetivo deste trabalho é obter uma ferramenta capaz de determinar o conjunto de variáveis mais adequado, tendo em consideração apenas a estrutura dos dados, conjunto de variáveis endógenas e exógenas, através da implementação de uma metodologia que proporcione índices que caracterizem a relação entre entradas e saídas de forma objetiva.

1.2 - Contribuições/Ferramentas

As principais contribuições para este trabalho foram:

- O conceito de Informação Mútua de Claude Shannon, que permite obter uma ideia sobre a relação e a partilha de informação entre duas variáveis;
- Neural network fitting tool (nftool), uma ferramenta da aplicação MATLAB que permite a construção e análise de redes neuronais artificiais, tendo em conta um conjunto de variáveis de entrada e de saída;
- Dados fornecidos pela empresa SmartWatt, contendo os valores horários de diversas variáveis relacionadas com Consumo e Produção Ibérica, incluindo variáveis meteorológicas, datando um período aproximado de três anos (33 meses): 33 variáveis cada uma com aproximadamente 24.139 linhas de dados.

1.3 - Estrutura da Dissertação

No capítulo 1 é apresentado o trabalho.

No capítulo 2 é apresentada a metodologia seguida durante o decorrer do trabalho.

No capítulo 3 são revistos alguns conceitos de entropia, e são apresentados alguns conceitos sobre Redes Neurais, aqueles com mais ênfase para este trabalho.

No capítulo 4 é aplicado o conceito de Informação Mútua ao nosso caso de estudo, onde se explicará os passos necessários para a aplicação do método, e se obterá os índices de IM para vários conjuntos de variáveis de entrada.

No capítulo 5 são colocados à prova os resultados obtidos pela aplicação do conceito de IM, através da construção e análise de RNs, com o auxílio da ferramenta nftool do MatLab. Os resultados serão testados tanto a nível individual, como conjuntos de variáveis de entrada.

As conclusões estarão no Capítulo 6, onde se também se encontra a síntese do método aplicado, juntamente com alguns cuidados a ter.

No Anexo A é apresentado o script de MatLab utilizado na obtenção dos valores dos índices de IM.

No Anexo B é apresentado o script de MatLab utilizado para a manipulação da variável Hora numa nova variável: Hora_168.

Por fim, no anexo C é apresentado o script de MatLab utilizado para a manipulação da variável DS, e criação da variável DS_8h.

Capítulo 2

Metodologia

Durante o desenvolvimento deste trabalho, a abordagem adotada foi desenvolvida com base nas seguintes preocupações:

- A aplicação dos métodos será realizada em problemas de previsão de carga, mas deverá ser possível estender os métodos a outros tipos de problemas que envolvam seleção de variáveis;
- As ações tomadas devem-se basear na análise da interdependência entre variáveis endógenas e exógenas, utilizando medidas objetivas (como a correlação, medidas de entropia, etc.);
- Os indicadores de qualidade para a seleção dos grupos de variáveis devem ser confirmados através de testes de desempenho.

Em conformidade com estes pontos, a metodologia adotada consiste nos seguintes passos principais:

1. Identificação de medidas que avaliem a relação entre a saída e cada uma das entradas;
2. Elaboração de indicadores/índices, com base nessas medidas, que permitam comparar o potencial de grupos de variáveis de entrada, em termos da qualidade relativa do desempenho de futuras previsões;
3. Realização de testes com diferentes conjuntos de entrada, e comparação da ordenação obtida com os indicadores do ponto anterior, com a ordenação dos desempenhos da previsão;
4. Síntese das principais conclusões, e estabelecimento do processo de seleção de variáveis.

Esta é a estrutura que foi adotada durante o desenvolvimento do trabalho realizado, tendo-se adotado a mesma estrutura na redação desta Dissertação.

Capítulo 3

Conceitos e Ferramentas

Neste capítulo são apresentadas duas noções fundamentais para o desenvolvimento e compreensão deste trabalho: o conceito de Entropia, e a ferramenta Rede Neuronal.

O conceito de entropia é indispensável para a definição do conceito de Informação Mútua, o método escolhido para auxiliar na seleção de variáveis, enquanto que as Redes Neurais foram a ferramenta de regressão utilizada nos testes de previsão que permitiu testar e verificar se os resultados do índice de IM são fidedignos.

3.1 - Entropia

O conceito de entropia, normalmente relacionado com a Termodinâmica[6], é um indicador da incerteza envolvida na previsão do valor de uma variável aleatória. De acordo com a 2ª Lei da Termodinâmica, a entropia de um sistema isolado nunca diminui, pois o sistema tende a evoluir para um estado de equilíbrio termodinâmico, um estado de entropia máxima.

O mesmo termo foi reutilizado por Claude Shannon, quando este apresentou o seu trabalho “A Mathematical Theory of Communication” em Julho e Outubro de 1948[10], onde Shannon estudou a estrutura de uma mensagem a ser transmitida por um canal com ruído como um processo estatístico, abrindo o caminho a ideias como a entropia da informação e informação mútua.

Um dos conceitos utilizados por Shannon foi introduzido por Hartley[7], em 1928, que refere que a quantidade de informação I para um dado evento x , que ocorre com uma probabilidade p , é dada pela equação $I = \log \frac{1}{p}$.

3.1.1 - Entropia e a teoria da informação

A teoria da informação diz-nos que a informação associada a um determinado evento é quanto maior quanto maior for a incerteza da ocorrência do evento. Em casos simples, a quantidade de informação pode ser obtida pelo logaritmo do número de casos possíveis[7].

Um exemplo muito simples são os casos do lançamento de uma moeda (dois resultados equiprováveis) e um dado (seis resultados equiprováveis). Saber o resultado do lançamento da moeda permite obter menos informação do que saber o resultado do lançamento do dado.

Claude Shannon, através do seu trabalho apresentado em 1948, apresentou a sua definição de entropia, como a quantidade de escolhas/decisões presentes num determinado evento, e a incerteza associada ao desenlace do mesmo, introduzindo os conceitos de entropia e informação mútua na área da teoria da informação.

Seja X um conjunto finito de variáveis, $X = \{x_1, x_2, x_3, \dots, x_n\}$, com probabilidades $p = \{p_1, p_2, p_3, \dots, p_n\}$, podemos definir a entropia (de Shannon) de X como[3]:

$$H(X) = - \sum_{i=1}^n p_i \log(p_i) \quad (\text{Eq. 3.1})$$

De notar que o cálculo da entropia não utiliza os valores de X , mas sim as suas probabilidades de ocorrência, apesar de ser utilizada a nomenclatura $H(X)$ em vez de $H(p)$.

Se X for um conjunto finito de variáveis aleatórias com uma distribuição de Bernoulli, onde a probabilidade do valor 0 é de $(1 - p)$, e do valor 1 é de p , a entropia de X é de:

$$H(X) = -p \log(p) - (1 - p) \log(1 - p) \quad (\text{Eq. 3.2})$$

Se a probabilidade de X for $p = 0$ ou $p = 1$ (X é determinístico), o valor da entropia é nulo, não existe incerteza. No entanto, se $p = 1/2$, caso da moeda por exemplo, o valor da entropia será máximo, o que indica um elevado grau de incerteza[5]. A figura seguinte representa estes resultados:

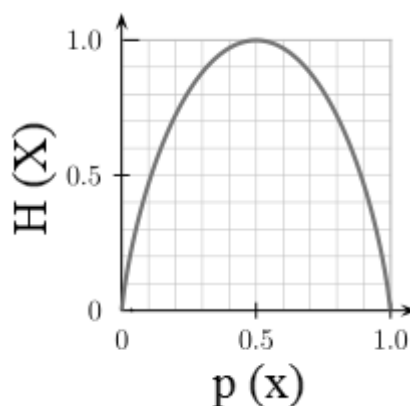


Figura 1 - Entropia esperada do lançamento de uma moeda justa

3.1.2 - Propriedades da Entropia

Algumas propriedades da entropia são[3]:

- $H(X) \geq 0$: A desigualdade mantém-se mesmo no caso de uma das probabilidades ser 1, e as restantes nulas;
- $H(X)$ é uma função contínua de p ;
- $H(X)$ é uma função simétrica, ou seja, a ordenação das probabilidades não influencia o valor de $H(X)$;
- A entropia de variáveis independentes é aditiva. Se X e Y são dois conjuntos de variáveis aleatórias independentes, com probabilidades p e q , então a entropia de (X,Y) será $H(X,Y) = H(X) + H(Y)$.

Estas propriedades são sempre válidas, independentemente do número de variáveis/conjuntos de variáveis a serem utilizadas. São também uma base para novos conceitos com fundamentos na Entropia, como os que veremos a seguir.

3.1.3 - Entropia conjunta e entropia condicional

Até aqui apenas se utilizou um único conjunto de variáveis para cálculos de entropia. Para casos de previsão, isso não é nem perto de suficiente, sendo necessário utilizar mais do que um conjunto de variáveis.

Consideremos dois conjuntos de variáveis aleatórias discretas, X e Y , com $X = \{x_1, x_2, x_3, \dots, x_n\}$ e $Y = \{y_1, y_2, y_3, \dots, y_m\}$, com probabilidades $p = \{p_1, p_2, p_3, \dots, p_n\}$ e $q = \{q_1, q_2, q_3, \dots, q_m\}$ respetivamente. Considerando ainda a probabilidade conjunta dos dois conjuntos de variáveis como $r(x_i, y_j)$ podemos obter, aplicando a expressão generalizada da entropia de Shannon para dois conjuntos de variáveis [3] [7]:

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m r(x_i, y_j) \log [r(x_i, y_j)], \quad (\text{Eq. 3.3})$$

assim como a expressão para a entropia condicional:

$$H(Y|x_i) = - \sum_{j=1}^m q(y_j|x_i) \log [q(y_j|x_i)], \quad (\text{Eq. 3.4})$$

onde $q(y_j|x_i)$ representa a probabilidade condicional de y_j , sabendo x_i .

Generalizando esta expressão para todo o conjunto X , multiplicando ambos os termos por $\sum_{i=1}^n p(x_i)$, obtemos:

$$H(Y|X) = - \sum_{i=1}^n \sum_{j=1}^m r(x_i, y_j) \log [q(y_j|x_i)], \quad (\text{Eq. 3.5})$$

Para estes dois novos casos, entropia conjunta e entropia condicional, as propriedades vistas anteriormente permanecem válidas, às quais se pode acrescentar:

- $H(Y|X) \leq H(Y)$, com a igualdade a acontecer se, e só se os conjuntos X e Y forem independentes.

Isto leva-nos a saber que informação acerca de X leva à diminuição da incerteza de Y . No caso de X e Y serem independentes, saber X não contribui para a diminuição da incerteza de Y . O mesmo é válido para $H(X|Y)$.

Analisando todas as propriedades vistas até aqui:

$$\begin{aligned} H(X, Y) &= H(X) + H(Y|X) \\ &= H(Y) + H(X|Y), \end{aligned} \quad (\text{Eq. 3.6})$$

podemos concluir que:

$$H(X, Y) = H(X) + H(Y|X) \leq H(X) + H(Y), \quad (\text{Eq. 3.7})$$

sendo a igualdade válida no caso de X e Y serem independentes.

3.1.4 - Informação Mútua

O conceito de Informação Mútua (I) é algo muito importante na teoria da informação e pode ser visto como a dependência entre duas variáveis aleatórias, ou como a quantidade de informação que uma variável contém sobre a outra, sendo definida por [2][3]:

$$\begin{aligned} I(X, Y) &= H(Y) - H(Y|X) \\ &= \sum_{i=1}^n \sum_{j=1}^m r(x_i, y_j) \log \frac{r(x_i, y_j)}{p(x_i) * q(y_j)}; \end{aligned} \quad (\text{Eq. 3.8})$$

$$= \int_Y \int_X r(x_i, y_j) \log \frac{r(x_i, y_j)}{p(x_i) * q(y_j)}, \quad (\text{Eq. 3.9})$$

em que $r(x_i, y_j)$ é a probabilidade conjunta de (x_i, y_j) . Estas duas expressões permitem saber a quantidade de informação que dois conjuntos partilham e o quanto saber um permite ajudar a determinar o outro.

Utilizando um diagrama de Venn, conseguimos chegar a uma ideia mais clara do que se pode obter utilizando a Informação Mútua:

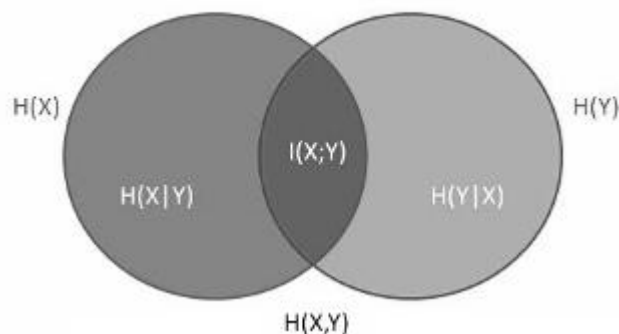


Figura 2 - Diagrama de Venn da Informação Mútua

A intersecção dos dois círculos não acontecerá caso os dois conjuntos sejam independentes, pois não existe informação conjunta aos dois grupos. Analisando o diagrama, consegue-se chegar a algumas das seguintes conclusões, e confirmar outras vistas já anteriormente[3]:

1. $H(X|Y) \leq H(X)$ e $H(Y|X) \leq H(Y)$;
2. $I(X,Y) \leq H(X), H(Y)$;
3. $I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$;
4. $H(X,Y) = H(X|Y) + I(X,Y) + H(Y|X) = H(Y) + H(X|Y) = H(X) + H(Y|X)$;
5. $H(X,Y) \leq H(X) + H(Y)$;

Assim, calculando os índices de Informação Mútua entre duas variáveis, uma exógena, a qual se quer determinar a sua importância, e outra endógena, sendo para o nosso caso de estudo a variável Consumo, conseguimos obter uma ideia do grau de importância e interligação entre essas variáveis, e a quantidade de informação que elas partilham.

3.2 - Redes Neurais

As redes neurais são modelos computacionais, inspirados na estrutura do nosso sistema nervoso, capazes de adquirir conhecimento através da experiência [20]. Através da utilização de um grande número de elementos computacionais (neurónios) interligados, sendo que a cada ligação, ou sinapse, está atribuído um peso, a rede tenta simular a capacidade do

cérebro humano de resolução de problemas. Estas redes são máquinas de aprendizagem, não lineares, e capazes de se adaptar [21].

Os neurónios, como se pode observar na figura 3, possuem três componentes básicos [23]:

- Um conjunto de sinapses provenientes da entrada, cada uma com um peso associado. O valor do peso pode ser qualquer valor positivo ou negativo;
- O somatório dos sinais de entrada, afetados pelos respectivos pesos;
- Uma função de ativação, responsável pela limitação do valor da saída do neurónio.

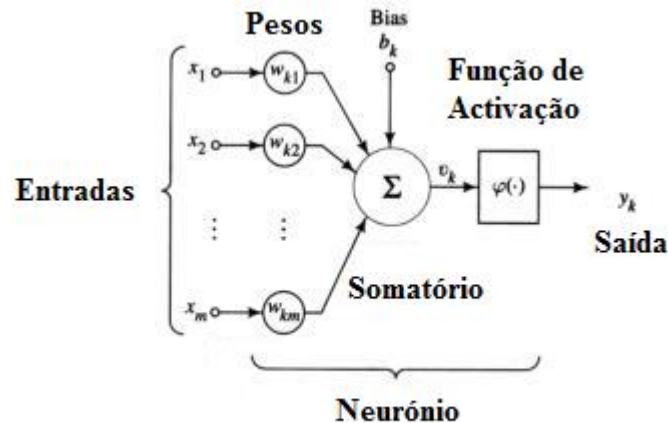


Figura 3 - Modelo não linear de um neurónio

As redes neuronais podem ser estáticas, quando a saída de cada neurónio depende apenas do valor atual da entrada, e dinâmicas se tiverem capacidade de memória de modo a aceder ao seu valor anterior [23].

Estas redes são constituídas por 3 elementos principais:

- Conjunto de entrada;
- Conjunto de saída;
- Conjunto de neurónios, também conhecida como a camada escondida. Os neurónios podem estar distribuídos por várias ou simplesmente uma única camada escondida.

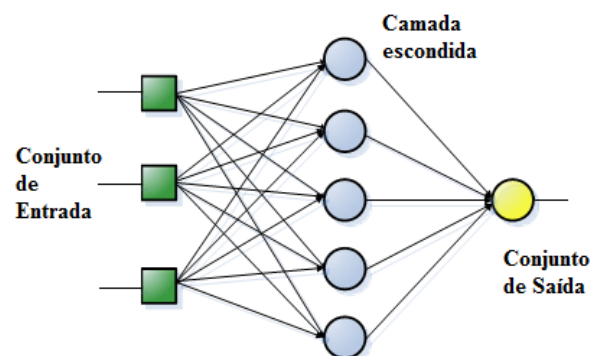


Figura 4 - Representação geral da estrutura de uma rede neuronal

As redes neurais são utilizadas para reconhecimento de padrões e/ou sequências, para processamento de dados ou sinais, entre outros usos, entre os quais, e mais importante para este trabalho, para previsão de séries temporais.

3.2.1 - Aprendizagem de uma rede neuronal

O treino de uma rede neuronal consiste na descoberta dos valores dos pesos que melhor relacionam os valores da entrada com os valores na saída. Cada passagem pela rede neuronal, durante a fase de treino, é considerada uma época. A prática normal é realizar um treino época a época, até os valores dos pesos estabilizarem, e o erro final convergir para um valor abaixo do limite estabelecido.

Uma quantidade de treino da rede excessiva pode levar à perda da capacidade de generalização. Este acontecimento é conhecido como overtraining [23].

3.2.1.1 - Aprendizagem por retro-propagação do erro

Este é o algoritmo usado pela ferramenta nftool do MatLab, que será utilizada mais tarde como ferramenta de teste para a verificação dos resultados.

A aprendizagem por retro-propagação do erro consiste num algoritmo que generaliza a aplicação do método dos mínimos quadrados a redes neurais, com mais de 2 camadas escondidas, que efetua o cálculo sucessivo das derivadas parciais na direção contrária à normal propagação da informação da rede, sendo esta aprendizagem do tipo supervisionada.

Numa aprendizagem supervisionada, conhece-se à partida os valores da entrada e saída pelo que se tenta construir um sistema mapeador capaz de reproduzir a função de transferência do sistema. O sistema produz uma resposta com base nos dados de entrada e, em resultado da comparação desse valor com o valor real, procede-se à modificação dos pesos das sinapses com base no erro verificado.

Para este tipo de aprendizagem, calcula-se o gradiente do erro em relação aos valores da camada de saída, obtendo-se assim o erro à saída da penúltima camada, continuando a propagação inversa ao sentido da informação sucessivamente até serem analisadas todas as camadas. Este processo permite atualizar os valores dos pesos, tudo com base no erro verificado à saída e é repetido até que este erro seja inferior à tolerância estabelecida.

3.2.2 - Vantagens e Desvantagens da utilização de redes neuronais

As RN permitem estabelecer sistemas de previsão, sem ser necessário conhecer a natureza da série de dados pois o processo de aprendizagem é capaz de adquirir as características essenciais do conjunto de dados, sendo capaz de trabalhar com funções lineares, e não lineares.

Uma outra grande vantagem das RNs é a sua capacidade de adaptação. Como podemos ver nos últimos dois capítulos, a rede neuronal consegue-se adaptar ao conjunto de dados, sendo possível treiná-la de forma a reduzir o erro.

No entanto, e de forma a podermos ser capazes de generalizar os resultados, é necessário dispor de um grande conjunto de dados para proceder ao processo de aprendizagem da rede. O tamanho da amostra deve exceder largamente o número de variáveis de entrada e saída.

De referir também que, à medida que a RN aumenta em complexidade, ou seja, que esta seja composta por um número muito grande de entradas e de camadas escondidas, o esforço computacional necessário é muito maior, o que leva a processos de aprendizagem muito mais lentos. No entanto, a sua habilidade de previsão continua a ser muito rápida, algo muito importante num sistema de previsão.

Um outro problema que pode surgir é a possibilidade da RN convergir para um mínimo local, em vez do mínimo global, o que leva a um erro significativo nas previsões. Como os pesos iniciais são inicializados aleatoriamente, retreinar a rede é possivelmente a melhor solução para este caso.

No geral, as redes neuronais são ferramentas muito flexíveis, que nos permitem trabalhar com sistemas complexos com relativa facilidade. As suas desvantagens, fora a quantidade de dados necessária para podermos obter bons resultados, podem ser facilmente ultrapassáveis.

Capítulo 4

Aplicação do Conceito de Informação Mútua a um caso prático

Dos métodos de entropia vistos anteriormente, a Informação Mútua é aquele surge como o mais indicado para a resolução do nosso problema de seleção de variáveis. Tratando-se de um problema de previsão de cargas, o método de informação Mútua permitir-nos-á medir a quantidade de informação que existe entre as várias variáveis de teste escolhidas, e a variável Consumo (MW), a variável para a qual se pretende seleccionar o melhor grupo de variáveis para a prever.

Seria de pensar que algo que já existe há algumas dezenas de anos já tenha sido minuciosamente estudado e compreendido. No entanto, desde o trabalho de Shannon em 1948, este assunto ainda se encontra a ser estudado. Devido ao facto de todas as expressões utilizarem não os valores dos conjuntos das variáveis, mas sim as suas probabilidades, dificulta a aplicação do método, tanto para uma distribuição discreta, como para uma distribuição contínua.

Para o caso em estudo, a seleção de variáveis para um problema de previsão de cargas, tendo em posse um conjunto de variáveis, e não a sua distribuição de probabilidades, é necessário adaptar as nossas variáveis ao método, de modo a que este seja possível de ser aplicado.

4.1 - Tratamento da Informação

A informação disponível, pertencente à empresa SmartWatt, é referente a um período de aproximadamente 3 anos de informação, ordenada cronologicamente. No total são 24.139 linhas de dados, com 33 variáveis distintas, sendo cada linha referente a uma hora do dia, estando incluídos nestes dados informações tais como o Consumo (MW), as produções Eólica,

Solar, Nuclear, entre outras; e dados meteorológicos (Temperatura, Pluviosidade, Direção e Velocidade do Vento, etc.). Estão consideradas as mudanças de hora realizadas ao longo do ano.

A estes dados foram aplicados alguns filtros, de modo a remover as linhas com campos em branco. Não houve necessidade de aplicar mais filtros, pois não haviam provas da existência de valores errôneos/exagerados. Isto resultou numa amostra final de 21.199 unidades, tendo sido estes os valores com que se desenvolveu o restante trabalho.

De todas as variáveis disponíveis, foram selecionadas algumas para se aplicar o conceito de Informação Mútua:

- Consumo (MW) - A nossa variável endógena, em que se pretende obter o melhor conjunto de variáveis para a sua previsão;
- DS - Dia da Semana, com Feriados. Assume uma gama de valores entre 1 e 8, um valor para cada dia da semana, com o 8 a corresponder aos dias de feriado;
- Hora - Hora do dia, outra variável discreta, com valores compreendidos entre 0 e 23;
- DAConsumo (MW) - Consumo registado 2 dias antes à mesma hora;
- SAConsumo (MW) - Consumo registado na semana anterior, à mesma hora;
- Temperatura ($^{\circ}\text{C}$) - Temperatura média durante aquela hora;
- Pluviosidade (mm^2) - Pluviosidade média Ibérica;
- Irradiância (W/m^2) - Irradiância média Ibérica;
- Vel. Vento (m/s) - Velocidade do vento média Ibérica.

Em previsões de carga para o dia seguinte, não é possível utilizar os valores do dia anterior ao dia para o qual se pretende prever, sendo necessário utilizar valores que os antecedam pelo menos 24 horas.

No caso da Temperatura, Pluviosidade, Irradiância e Vel. Vento, tratam-se de valores reais, medidos e registados, algo que, durante previsões, o previsor não tem acesso, tendo que se sujeitar a utilizar ou previsões destas variáveis, ou usar aproximações, com base em valores que se registaram anteriormente para cada variável.

Como o tamanho da nossa amostra é muito superior à quantidade de variáveis escolhidas, não haverá problemas de generalização.

Como já foi referido anteriormente, a expressão para o cálculo da IM utiliza probabilidades, em vez dos valores em si, sendo necessário proceder a uma transformação da informação em posse.

A solução para ser possível implementar a Informação Mútua passa por transformar as nossas variáveis numa distribuição de probabilidades, definindo a sua dimensão e o passo entre “escalões” diferentes.

4.1.1 - Construção da matriz das probabilidades

A matriz das probabilidades é única para cada combinação de variáveis, podendo calculada com auxílio de ferramentas capazes de trabalhar com matrizes, tendo sido utilizada, para este caso, o MatLab.

Para tal foi desenvolvido um script onde seria calculada os valores da matriz, sendo apenas necessário indicar as variáveis a serem utilizadas, o número de linhas e colunas (dimensão) e o passo entre linhas e colunas que se pretende ter na matriz final [9].

Tomando como exemplo o cálculo desta matriz para as variáveis Dia da Semana (DS) e Consumo. A variável DS encontra-se limitada entre 1 e 8, sendo que o valor 1 corresponde aos Domingos, os valores 2 a 6 correspondem aos dias da semana (2ª a 6ª feira respetivamente), o valor 7 corresponde aos Sábados, e por fim, o valor 8 corresponde aos dias de Feriado. Tratando-se de uma variável discreta (variável que assume uma gama pequena de valores), considera-se um passo unitário entre as linhas da matriz.

A variável Consumo encontra-se na gama de valores [21.470; 52.853]. O passo a ser considerado está dependente do número de colunas. Para este exemplo, pretende-se construir uma matriz quadrada, sendo que ao contrário da variável DS, o número de divisões da variável Consumo não está limitado. Assim, para 8 colunas, o passo será de $(\text{Máx} - \text{Min}) / 8 = 3922$. Isto significa que cada coluna considerará uma gama de valores igual a 3922 MW.

Visto isto, e aplicando o script de MatLab para estas duas variáveis, para os valores/dimensões definidas, obtemos a seguinte matriz de probabilidades para as variáveis DS vs. Consumo:

		Consumo								Soma
		21470-25392	25393-29315	29316-33237	33238-37160	37161-41083	41084-45006	45007-48929	48930-52853	
D S	1	2,245%	4,326%	4,783%	1,604%	0,542%	0,160%	0,005%	0,000%	13,666%
	2	0,986%	2,415%	1,283%	2,373%	3,901%	1,929%	0,675%	0,118%	13,680%
	3	0,094%	2,283%	1,868%	2,113%	4,297%	2,467%	0,807%	0,090%	14,020%
	4	0,113%	2,094%	1,967%	2,203%	4,161%	2,557%	0,741%	0,085%	13,920%
	5	0,090%	1,967%	2,024%	2,033%	4,269%	2,293%	0,679%	0,118%	13,472%
	6	0,066%	1,882%	1,977%	2,491%	4,113%	1,925%	0,609%	0,071%	13,133%
	7	0,316%	3,109%	4,439%	4,401%	1,109%	0,283%	0,042%	0,000%	13,699%
	8	0,627%	1,208%	1,208%	0,844%	0,443%	0,047%	0,028%	0,005%	4,411%
Soma		4,538%	19,284%	19,548%	18,062%	22,836%	11,661%	3,585%	0,486%	100,000%

Tabela 1- Matriz Probabilidades DS vs Consumo (em %)

Através desta tabela, é possível obter todos os elementos necessários para aplicar a expressão de cálculo da Informação Mútua, pois possuímos os valores das probabilidades conjuntas, os valores no interior das células, e os valores das probabilidades de cada variável, dados pelos somatórios segundo o índice de linha (para a variável DS) e segundo o índice das

colunas (para a variável Consumo). De acrescentar que a matriz serve apenas para o cálculo dos índices de IM, e não como um método alternativo de representar o conjunto de dados. Na fase de testes serão usados os valores reais, e não os valores obtidos para estas matrizes.

4.2 - Cálculo dos índices de Informação Mútua

O cálculo dos índices de Informação Mútua é uma simples aplicação da expressão para valores discretos, tendo como base as matrizes calculadas através do que foi visto no ponto anterior, para cada variável em análise. Relembrar que as matrizes são sempre construídas entre uma variável de teste, e a variável Consumo, a variável para qual pretendemos determinar o melhor conjunto de variáveis para a prever.

Depois de selecionar algumas variáveis para o conjunto de teste, e analisar as dimensões possíveis para as matrizes de probabilidades, procedeu-se aos cálculos dos índices de Informação Mútua. Como regra geral, e tendo em atenção que as variáveis discretas estão limitadas no que toca a possíveis dimensões das matrizes de probabilidades, estas serão sempre matrizes quadradas.

Para as dimensões (8 x 8) e (24 x 24), obteve-se os seguintes valores de Informação Mútua:

Variavel	Dimensão	
	8	24
DS	0,082	
Hora		0,246
DaConsumo	0,127	0,165
SaConsumo	0,363	0,458
Temperatura	0,057	0,077
Pluviosidade	0,002	0,007
Vel. Vento	0,005	0,013
Irradiância	0,034	0,046
Soma	0,6699	1,0107

Tabela 2 - Coeficientes de IM para dimensões 8 e 24

Variavel	Dimensão	
	8	24
DS	12,27%	
Hora		24,29%
DaConsumo	18,93%	16,30%
SaConsumo	54,25%	45,30%
Temperatura	8,45%	7,66%
Pluviosidade	0,22%	0,66%
Vel. Vento	0,78%	1,26%
Irradiância	5,11%	4,54%
Soma	100%	100%

Tabela 3 - Importância relativa das variáveis de acordo com os índices de IM para dimensões 8 e 24

Apesar de se poder tirar algumas conclusões das tabelas anteriores, é arriscado estar a apressar sem estar na posse de mais informações. No entanto, é possível verificar o peso, em ambas as tabelas de algumas variáveis.

Como não é possível trabalhar mais com as variáveis Hora e DS, devido às suas limitações, o próximo passo é descartá-las, temporariamente, e proceder ao cálculo dos índices de IM para diversas dimensões, de modo a poder observar a sua evolução com a dimensão das matrizes.

4.3 - Evolução do índice de Informação Mútua

Ignorando as variáveis Hora e DS, pois estas limitam as dimensões que se pode utilizar, é possível calcular os índices de IM para diversas dimensões, e assim analisar a sua evolução:

	Dimensão									
Variavel	8	24	25	50	75	100	150	168	200	250
DaConsumo	0,127	0,165	0,167	0,190	0,214	0,243	0,314	0,344	0,404	0,492
SaConsumo	0,363	0,458	0,461	0,489	0,506	0,528	0,575	0,591	0,629	0,683
Temperatura	0,057	0,077	0,078	0,093	0,117	0,147	0,225	0,257	0,320	0,418
Pluviosidade	0,002	0,007	0,007	0,019	0,035	0,055	0,097	0,113	0,144	0,193
Vel. Vento	0,005	0,013	0,014	0,030	0,055	0,085	0,159	0,187	0,245	0,337
Irradiância	0,034	0,046	0,045	0,064	0,092	0,128	0,211	0,241	0,294	0,371
Soma	0,588	0,765	0,7714	0,885	1,019	1,1852	1,582	1,733	2,0357	2,495

Tabela 4 - Evolução dos índices de IM de acordo com a dimensão da matriz das probabilidades

	Dimensão									
Variavel	8	24	25	50	75	100	150	168	200	250
DaConsumo	21,58%	21,52%	21,60%	21,44%	21,03%	20,54%	19,84%	19,85%	19,84%	19,70%
SaConsumo	61,83%	59,83%	59,76%	55,19%	49,69%	44,56%	36,36%	34,11%	30,89%	27,39%
Temperatura	9,63%	10,12%	10,10%	10,52%	11,43%	12,36%	14,23%	14,81%	15,71%	16,77%
Pluviosidade	0,26%	0,88%	0,95%	2,19%	3,47%	4,63%	6,13%	6,52%	7,05%	7,73%
Vel. Vento	0,88%	1,66%	1,79%	3,43%	5,39%	7,14%	10,08%	10,80%	12,04%	13,52%
Irradiância	5,82%	6,00%	5,81%	7,23%	8,98%	10,77%	13,36%	13,90%	14,46%	14,89%
Soma	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Tabela 5 - Evolução da importância relativa das variáveis de acordo com os índices de IM

Estes valores foram obtidos utilizando matrizes quadradas, com as dimensões definidas no índice de coluna.

Com base nestes dados, torna-se possível tirar algumas conclusões:

1. Os valores dos coeficientes de IM aumentam à medida que as dimensões da matriz de probabilidades também aumentam;
2. As importâncias relativas, regra geral, sofrem alterações à medida que se aumenta as dimensões da matriz (a única variável que se mantém relativamente constante é a variável DAConsumo);

3. Apesar da variação das importâncias relativas, a ordenação das mesmas (da maior para a menor, ou vice-versa) é sempre a mesma, independentemente da dimensão da matriz das probabilidades;

Como é possível observar, e de acordo com o ponto 1 acima, o índice de IM varia de acordo com as dimensões da matriz. Esta evolução é expectável, se aplicarmos os conceitos vistos anteriormente de Entropia e Informação Mútua.

Graficamente, é possível obter uma melhor percepção desta evolução, para as várias variáveis:

DAConsumo	
Dimensões	IM
250/250	0,4916
200/200	0,4038
168/168	0,3441
150/150	0,3138
100/100	0,2434
75/75	0,2143
50/50	0,1898
25/25	0,1666
24/24	0,1647
8/8	0,1268

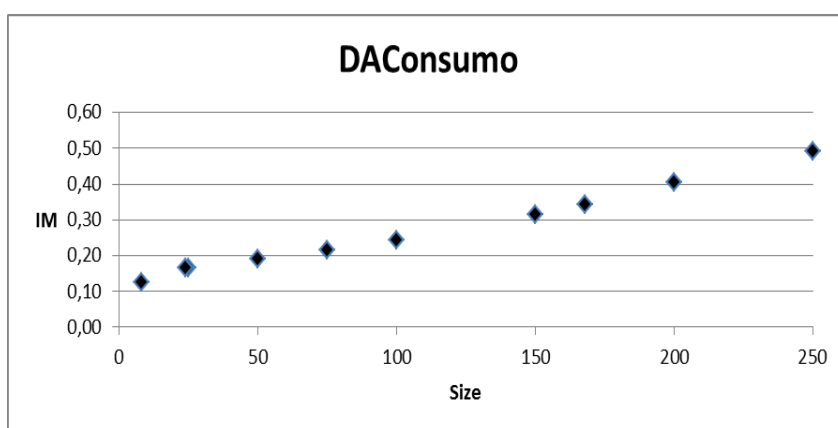


Tabela 6 - Evolução dos valores da variável DAConsumo com a dimensão da matriz das probabilidades

Figura 5 - Representação gráfica da evolução dos valores da variável DAConsumo com a dimensão da matriz das probabilidades

SAConsumo	
Dimensões	IM
250/250	0,6833
200/200	0,6289
168/168	0,5913
150/150	0,5751
100/100	0,5281
75/75	0,5063
50/50	0,4886
25/25	0,461
24/24	0,4578
8/8	0,3634

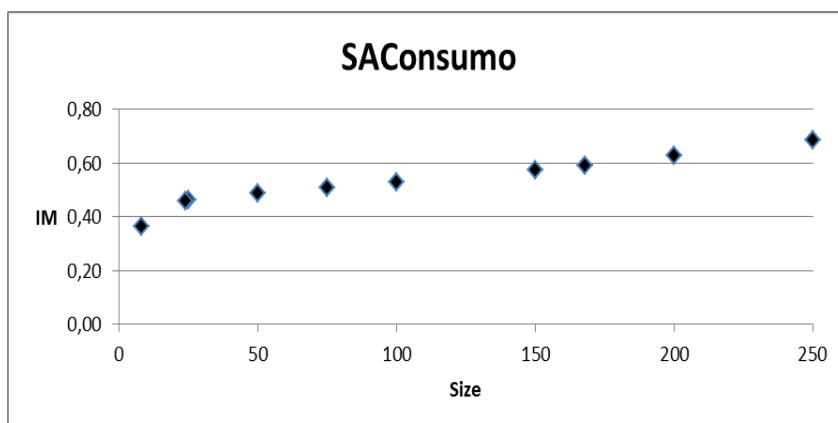


Tabela 7 - Evolução dos valores da variável SAConsumo com a dimensão da matriz das probabilidades

Figura 6 - Representação gráfica da evolução dos valores da variável SAConsumo com a dimensão da matriz das probabilidades

Temperatura	
Dimensões	IM
250/250	0,4184
200/200	0,3199
168/168	0,2568
150/150	0,2251
100/100	0,1465
75/75	0,1165
50/50	0,0931
25/25	0,0779
24/24	0,0774
8/8	0,0566

Tabela 8 - Evolução dos valores da variável Temperatura com a dimensão da matriz das probabilidades

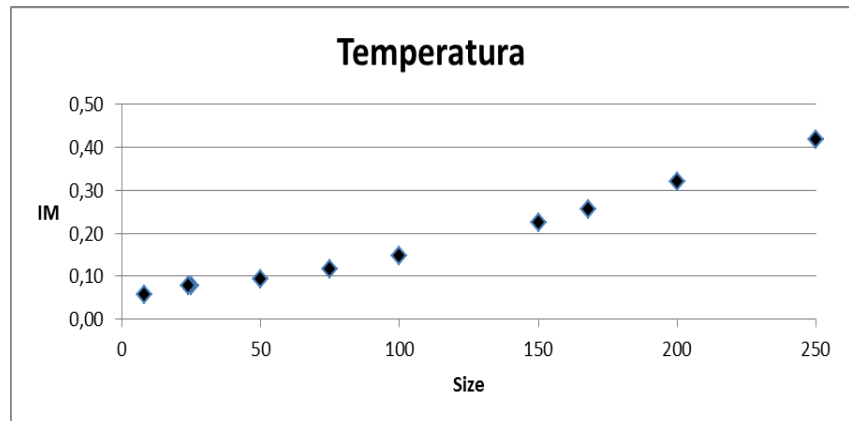


Figura 7 - Representação gráfica da evolução dos valores da variável Temperatura com a dimensão da matriz das probabilidades

Pluviosidade	
Dimensões	IM
250/250	0,1928
200/200	0,1435
168/168	0,113
150/150	0,097
100/100	0,0549
75/75	0,0354
50/50	0,0194
25/25	0,0073
24/24	0,0067
8/8	0,0015

Tabela 9 - Evolução dos valores da variável Pluviosidade com a dimensão da matriz das probabilidades

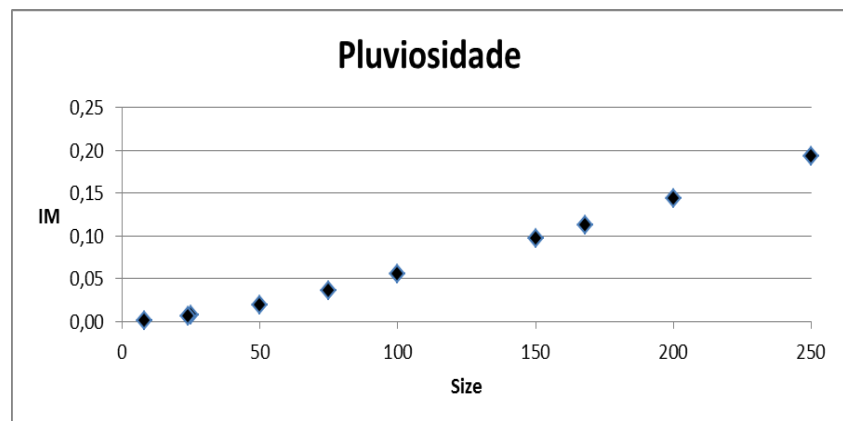


Figura 8 - Representação gráfica da evolução dos valores da variável Pluviosidade com a dimensão da matriz das probabilidades

Vel. Vento	
Dimensões	IM
250/250	0,3373
200/200	0,2452
168/168	0,1872
150/150	0,1594
100/100	0,0846
75/75	0,0549
50/50	0,0304
25/25	0,0138
24/24	0,0127
8/8	0,0052

Tabela 10 - Evolução dos valores da variável Vel. Vento com a dimensão da matriz das probabilidades

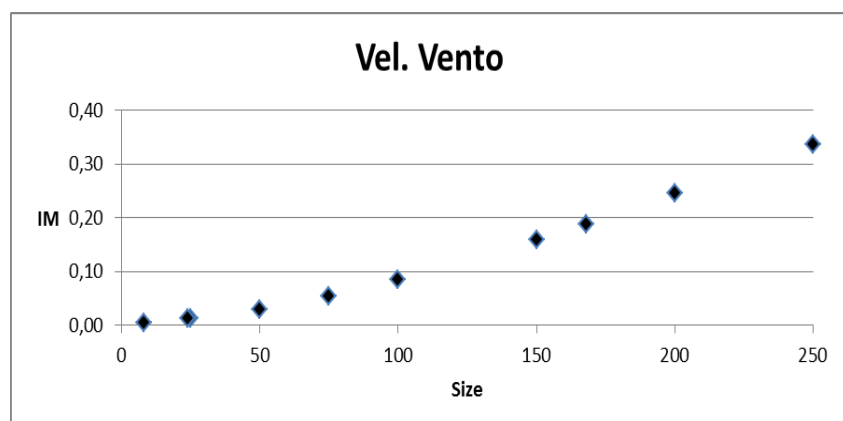


Figura 9 - Representação gráfica da evolução dos valores da variável Vel. Vento com a dimensão da matriz das probabilidades

Irradiância	
Dimensões	IM
250/250	0,3714
200/200	0,2944
168/168	0,241
150/150	0,2113
100/100	0,1277
75/75	0,0915
50/50	0,064
25/25	0,0448
24/24	0,0459
8/8	0,0342

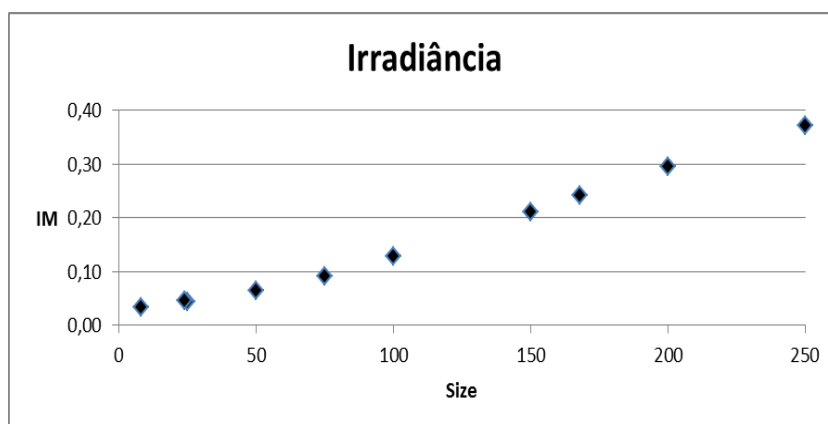


Figura 10 - Representação gráfica da evolução dos valores da variável Irradiância com a dimensão da matriz das probabilidades

Tabela 11 - Evolução dos valores da variável Irradiância com a dimensão da matriz das probabilidades

Como se pode observar pelos gráficos anteriores, os índices de IM têm uma evolução quase linear, à medida que se aumenta as dimensões da matriz de probabilidades usada no seu cálculo.

O aumento das dimensões da matriz implica um aumento da incerteza do sistema (aumento da Entropia), o que leva a um aumento da Informação Mútua. Com base nisto, decidiu-se utilizar importâncias relativas, em vez de tirar conclusões com base nos valores. Apesar das importâncias relativas também variarem com a variação das dimensões da matriz das probabilidades, como foi referido no ponto 3 acima, a ordenação das variáveis mais importantes/influentes é inalterável.

De referir que, devido aos processos matemáticos nos cálculos intermédios, o tempo de cálculo dos índices associados a dimensões mais pequenas, é muito inferior ao tempo de cálculo de índices associados a matrizes de dimensões maiores (o tempo de cálculo de índices de IM com matrizes de probabilidade de (200 x 200) ou (250 x 250) era superior a 30 min, para cada variável, enquanto que para dimensões menores, (8 x 8) ou (24 x 24) por exemplo, os resultados eram obtidos em apenas alguns segundos). A isto acrescenta-se que a dimensão que for escolhida, fora as variáveis discretas, irá afetar o passo, ou seja, a gama de valores em cada linha/coluna. Para dimensões muito grande, isto resultará numa grande fragmentação da gama de valores, resultando na possibilidade de uma grande quantidade de células com probabilidade nula. Para dimensões demasiado pequenas, não haverá seletividade, e a ferramenta probabilidade perde a sua função.

4.4 - Manipulação das variáveis discretas

Como foi possível observar no capítulo 3.2, as variáveis discretas estão muito limitadas no que se refere à utilização da Informação Mútua. No nosso caso, tanto a variável Hora, como a variável DS só podem construir uma matriz com uma única dimensão diferente uma da outra. Como já foi analisado, os índices de IM evoluem consoante as dimensões da matriz das probabilidades o que, para este caso, não nos permite mensurar a importância destas duas variáveis. Individualmente, para cada dimensão, há indicações que são variáveis com alguma importância, especialmente a variável Hora, para a previsão do Consumo, mas neste momento não é possível dizer com certeza.

Dado este problema, surgem algumas soluções:

1. Estimar o valor para uma nova dimensão da MP;
2. Manipular a variável, de modo a que esta seja capaz de produzir uma matriz com novas dimensões;

Como conseguimos ver nos gráficos e tabelas do ponto anterior, cada variável tem a sua própria evolução. Para dificultar ainda mais, apenas possuímos um único valor de cada variável, o que dificulta a sua previsão para novas dimensões.

A solução encontrada foi a manipular estas variáveis, de modo a conseguirmos construir uma matriz de probabilidades de maior dimensão, podendo ser necessário recorrer a informação contida noutras variáveis.

4.4.1 - Variável Hora_168

Tendo em posse informação sobre ambos o Dia da Semana e Hora, é possível obter uma variável que represente a hora semanal. Uma semana tem 168 horas e, após escolher a hora zero, com base nas variáveis Hora e DS pode-se construir uma nova variável (Anexo X). Neste caso, os dias classificados como Feriados são tratados como Domingos, devido a semelhanças nos consumos registados nesses dias.

Para esta nova variável, calculou-se o seu índice de IM, sendo depois possível compará-la com as restantes variáveis para a mesma dimensão:

	Dimensão
Variavel	168
DaConsumo	0,344
SaConsumo	0,591
Temperatura	0,257
Pluviosidade	0,113
Vel. Vento	0,187
Irradiância	0,241
Hora_168	0,450
Soma	2,183

Tabela 12 - Coeficientes de IM para a dimensão 168

	Dimensão
Variavel	168
DaConsumo	15,76%
SaConsumo	27,09%
Temperatura	11,76%
Pluviosidade	5,18%
Vel. Vento	8,58%
Irradiância	11,04%
Hora_168	20,59%
Soma	100,00%

Tabela 13 - Importância relativa das variáveis de acordo com os índices de IM para a dimensão 168

A nova variável apresenta resultados prometedores para a fase de teste, e permite a comparação com as restantes variáveis, ao contrário do que sucedia com as variáveis originais. No entanto, dois problemas surgem:

1. A nova variável continua a não permitir alterações nas dimensões da matriz das probabilidades;
2. Esta nova variável será capaz de produzir melhores, piores, ou será igual se for utilizada num cenário de previsão vs. as duas variáveis originais?

Em relação ao primeiro ponto, não há alternativa. Apesar da alteração efetuada, a nova variável continua limitada a uma única dimensão da sua MP. O segundo ponto será visto mais tarde, quando se efetuar os testes das variáveis através da utilização de redes neuronais.

Através da observação dos valores obtidos, consegue-se facilmente ver que esta variável apresenta um índice muito elevado, sendo a segunda variável com o índice mais alto, o que demonstra a importância que as variáveis DS e Hora têm na previsão de carga.

4.4.2 - Variável DS_8h

Aproveitando o facto de a variável estar dimensionada para 8 valores distintos, em vez de apenas para 7, devido à presença de um valor distinto para os feriados, é possível, com ajuda da variável Hora, criar uma nova variável com dimensão 24, o que permitiria a comparação posterior de valores para a mesma dimensão (Ver anexo Y).

Fragmentando o dia em compassos de 8 horas, obtemos uma nova variável que assume 24 valores distintos. No entanto, e tal como acontece no caso anterior, o facto de utilizarmos informações de outras variáveis, faz-nos olhar com outros olhos para os resultados finais pois nestes estão refletidos informação de mais do que uma variável.

Relativamente a este caso, o acréscimo de informação que a variável Hora vem dar não é muito significativo, visto que esta apenas é usada como guia para a segmentação em períodos de 8 horas, como é possível ver no script de MatLab.

Com a nova variável obtiveram-se os seguintes valores:

	Dimensão
Variavel	24
DS_8h	0,265
Hora	0,246
DaConsumo	0,165
SaConsumo	0,458
Temperatura	0,077
Pluviosidade	0,007
Vel. Vento	0,013
Irradiância	0,046
Soma	1,2755

Tabela 14 - Coeficientes de IM para a dimensão 24 c/ a variável DS_8h

	Dimensão
Variavel	24
DS	20,76%
Hora_8h	19,25%
DaConsumo	12,91%
SaConsumo	35,89%
Temperatura	6,07%
Pluviosidade	0,53%
Vel. Vento	1,00%
Irradiância	3,60%
Soma	100%

Tabela 15 - Importância relativa das variáveis de acordo com os índices de IM para a dimensão 24 c/ a variável DS_8h

Sendo agora possível fazer uma comparação, é possível verificar a importância das duas variáveis discretas, apesar da variável SAConsumo apresentar um índice que sugere ser a variável que trará melhores resultados.

4.5 - Variáveis Meteorológicas

Da lista de variáveis selecionadas constam 4 variáveis meteorológicas, sendo elas:

- Temperatura ($^{\circ}\text{C}$) - Temperatura média durante aquela hora;
- Pluviosidade (mm^2) - Pluviosidade média Ibérica;
- Irradiância (W/m^2) - Irradiância média Ibérica;
- Vel. Vento (m/s) - Velocidade do vento média Ibérica.

A particularidade destas variáveis está relacionada com a sua disponibilidade na altura de efetuar a previsão. Os dados em posse são referentes aos valores reais registados hora a hora. Numa situação de previsão de cargas, o previsor não tem acesso a estes dados, tendo em vez disso acesso ou a previsões dessas variáveis, ou ter que utilizar valores antigos, sabendo que estes serão relativamente próximos dos valores que irão surgir.

Assim sendo, não é completamente correto determinar os índices de Informação Mútua para estas variáveis, sabendo que os valores disponíveis na altura da previsão não seriam estes.

Para tal, procedeu-se à determinação dos índices de IM destas variáveis, mas referentes ao dia anterior, de modo a verificar o impacto de usar uma ou outra variável:

Variavel	Dimensão	
	24	100
DATemperatura	0,0797	0,1487
DAPluviosidade	0,0074	0,0545
DAVel. Vento	0,0123	0,0838
DAIrradiância	0,0490	0,1337
Soma	0,1484	0,4207

Tabela 16 - Índices de IM das variáveis meteorológicas

Variavel	Dimensão	
	24	100
Temperatura	0,0771	0,1471
Pluviosidade	0,0067	0,0555
Vel. Vento	0,0134	0,0866
Irradiância	0,0461	0,1293
Soma	0,1433	0,4185

Tabela 17 - Índices de IM das variáveis meteorológicas medidas 2 dias antes

À exceção da variável DAVel. Vento, todas as variáveis DA apresentam melhores resultados que as variáveis que referem os valores verificados em tempo real para pelo menos uma dimensão da matriz das probabilidades. Os valores são muito próximos no entanto, o que sugere que segundo a Informação Mútua, não existem diferenças significativas em usar uma ou outra variável. O impacto de usar uma ou outra variável terá que ser determinado através dos resultados dos testes das redes neuronais que serão vistos mais à frente.

Estando todos os índices necessários calculados, é preciso verificar se existe uma relação entre os mesmos e o erro produzido pela utilização das variáveis. Para tal vai-se proceder à construção e análise de redes neuronais, num primeiro passo com apenas 1 variável, e mais tarde alargando o número de entradas.

Capítulo 5

Testes de Desempenho

Com a obtenção de todos os índices de IM necessários, torna-se necessário verificar o quão credível são estes valores, e se o conceito de IM é algo que se pode aplicar num cenário de previsão de cargas. Para tal serão usadas redes neuronais para construir e analisar as redes neuronais obtidas utilizando as variáveis do grupo de teste já indicado.

Inicialmente, as redes neuronais serão construídas na base 1:1 (1 entrada para 1 saída). Apesar de esta não ser um procedimento normal, pois numa previsão, não seria utilizada apenas uma variável, isto permitir-nos-á ter uma ideia sobre a ordem de grandeza do erro, e se existe uma correlação entre esse mesmo erro, e os índices obtidos anteriormente.

Posteriormente, serão usadas mais entradas, de modo a verificar se existe uma redução do erro consistente com o aumento dos índices das variáveis utilizadas, assim como tentar perceber se e como as variáveis se relacionam entre si.

De modo a mensurar o erro obtido pelas redes neuronais, serão utilizadas duas medidas: o MSE (Mean Square Error) e o MAPE (Mean Absolute Percentage Error). O primeiro é a medida utilizada pelo nftool, a ferramenta do MatLab que permite a criação e estudo das redes neuronais, como critério de minimização. O MAPE, devido à sua apresentação em percentagem, irá permitir estabelecer melhor as ligações entre os erros das redes neuronais, e o valor total dos índices de IM das variáveis utilizadas. Apesar de serem 2 formas de mensurar o erro, estas duas medidas não se comportam obrigatoriamente da mesma forma, ou seja, a ordenação das variáveis em relação ao MSE, poderá não ser a mesma quando considerado o MAPE.

5.1 - Criação e análise de redes com uma variável de entrada

Após transferir todas as variáveis para a área de trabalho, procedeu-se à construção das redes neuronais. Para cada um dos casos, o output (a variável Consumo) e o número de

neurónios da camada escondida, definido a 10 para este caso, não foram alterados, permitindo ter condições de igualdade para todas as variáveis.

Os resultados obtidos foram os seguintes:

Variavel	MAPE	MSE
DS	13,39%	29462113,11
Hora	9,24%	17140302,57
DaConsumo	11,13%	22034528,71
SaConsumo	4,42%	5088724,84
Temperatura	14,39%	32207474,52
Pluviosidade	15,47%	36192404,72
Vel. Vento	15,44%	35898305,09
Irradiância	14,32%	32580220,99
Soma		

Tabela 18 - Valores do MAPE e MSE das redes neuronais c/ 1 entrada

Recordando que, devido às características das variáveis discretas, não foi possível obter os índices de todas as variáveis para uma dimensão da MP em comum a todas. Assim, a comparação dos erros com os índices será feita considerando uma dimensão de 24 para todas as variáveis, exceto para a variável DS, onde será considerada o índice obtido com a matriz de dimensão 8.

Variavel	MAPE	MSE	MI	Razão	Dimensão
DS	13,39%	29462113	0,0822	7,52%	8
Hora	9,24%	17140303	0,2455	22,46%	24
DaConsumo	11,13%	22034529	0,1647	15,07%	24
SaConsumo	4,42%	5088725	0,4578	41,89%	24
Temperatura	14,39%	32207475	0,0774	7,08%	24
Pluviosidade	15,47%	36192405	0,0067	0,61%	24
Vel. Vento	15,44%	35898305	0,0127	1,16%	24
Irradiância	14,32%	32580221	0,0459	4,20%	24
Soma			1,0929	100%	

Tabela 19 - Comparação entre os índices de IM, e os erros obtidos pelas redes neuronais

Analisemos os resultados obtidos:

- Ordenando as variáveis por ordem crescente do MSE, obtemos uma ordenação decrescente do índice de IM;

- É também possível observar que a ordenação segundo o MSE não é exatamente igual à ordenação considerando o MAPE. No entanto, ambos os critérios apresentam evoluções semelhantes quando equiparados ao índice MI;
- A variável DS, mesmo sendo comparada com um índice calculado com uma matriz de menor dimensão, apresenta-se como uma das variáveis mais influentes (4ª variável com melhor resultado).

A primeira conclusão é bastante importante. Segundo estes resultados, as variáveis com os índices de Informação Mútua mais elevados são as que apresentam os melhores resultados, aqueles com o menor erro. Isso é possível confirmar através da seguinte representação gráfica do MSE em função da razão IM:

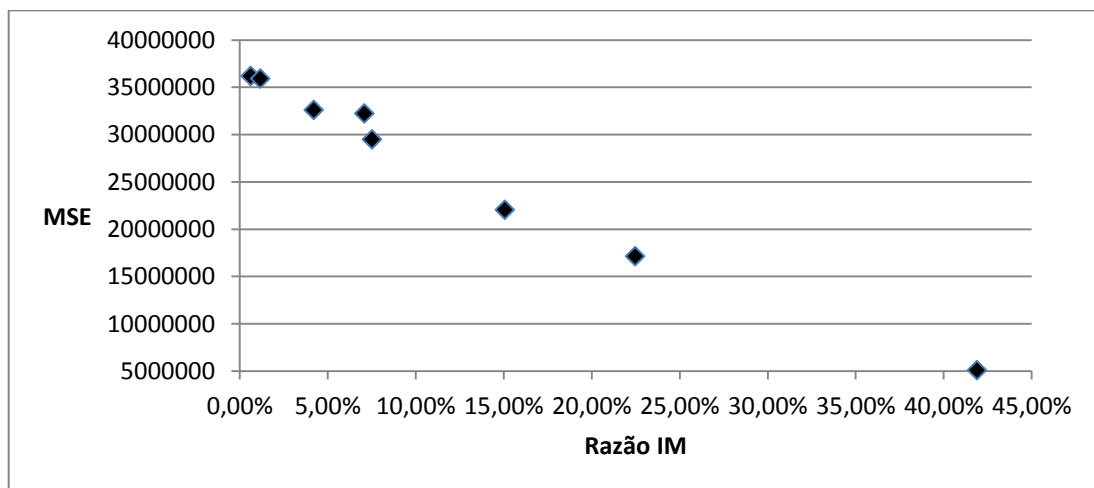


Figura 11 - Representação gráfica do MSE em função da razão IM, para redes neurais com 1 entrada

Graças à figura 11 representada acima, é facilmente compreensível que o aumento do índice de Informação Mútua leva a uma diminuição do erro quadrático.

No entanto, e tal como foi dito no início, apesar de esta ter sido uma forma de relacionar o erro dos resultados obtidos das redes neurais com o índice e IM, numa situação de previsão não será apenas utilizada uma variável, pelo que será necessário estudar o que sucede com a introdução de mais variáveis ao conjunto de entrada.

5.2 - Criação e análise de redes com múltiplas entradas

Não estando agora as variáveis de entrada limitadas a um conjunto único, o número de combinações de variáveis é imenso. Não sendo possível vê-las todas, analisemos alguns dos pontos mais importantes.

Muito foi discutido acerca das limitações das variáveis discretas Dia da Semana e Hora, mas tal como foi possível observar na Tabela 19, elas são 2 das variáveis mais influentes. Mas será o suficiente para, quando utilizadas em conjunto, serem capazes de uma previsão eficaz?

Utilizando as duas variáveis, unicamente, e em conjunto com outras, obteve-se os seguintes resultados:

Variáveis	MAPE	MSE	Razão MI
Hora,DS	6,06%	7758647	29,98%
Hora,DS,SAConsumo	3,96%	3580659	71,87%
Hora,DS,DAConsumo	3,32%	2604014	45,05%
Hora,DS,DAConsumo,SAConsumo	2,85%	1909901	86,94%
Hora, DS, DAConsumo,SAConsumo,Temperatura	2,63%	1624445	94,03%
Hora, DS, DAConsumo,SAConsumo,Temperatura,Irradiância	2,54%	1515141	98,22%
Hora,DS,Temperatura	4,33%	4025539	37,07%
Hora,DS,Temperatura,Pluviosidade	4,29%	3942885	37,68%
Hora, DS, Daconsumo, Temperatura	3,47%	2589453	52,14%
Hora, DS, SAConsumo, Temperatura	3,60%	2942514	78,96%

Tabela 20 - Comparação entre os índices de IM e o MSE c/ DS e Hora no conjunto de entradas

De modo a facilitar a interpretação dos dados, a figura 12 representa os dados da tabela anterior graficamente:

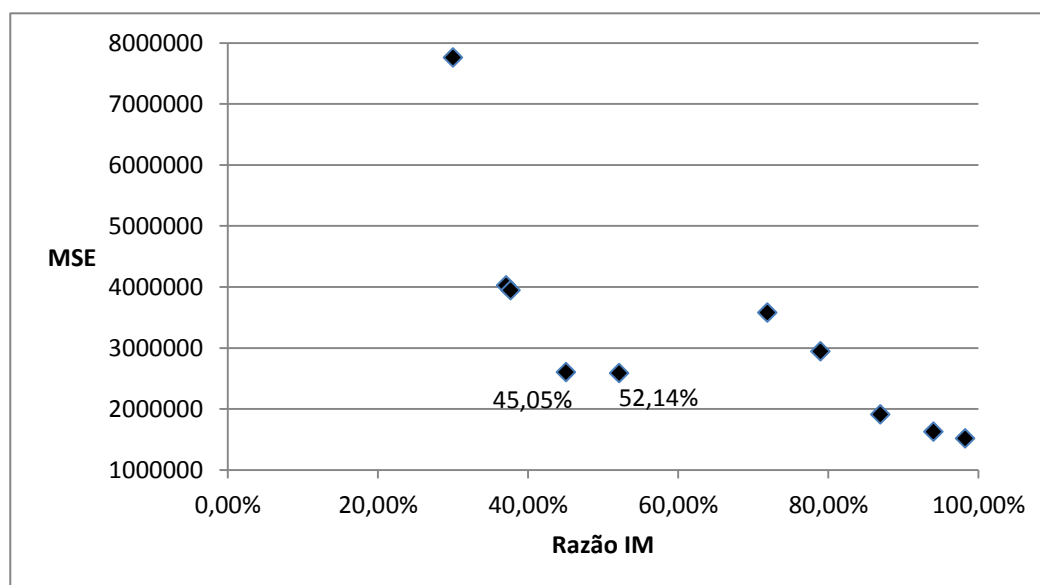


Figura 12 - Representação gráfica do MSE em função da razão IM, para redes neurais c/ DS e Hora no conjunto de entradas

Apesar de estarem na 2ª e 4ª posição respectivamente no que toca a resultados individuais, as variáveis Hora e DS perfazem uma percentagem muito baixa quando utilizadas juntas (considerando os índices de IM), e apresentam um erro bastante elevado. Apenas quando surge uma terceira variável no conjunto das variáveis de entrada é que o valor do erro desce, e substancialmente.

Consegue-se reparar também que a adição de variáveis ao conjunto de entrada tem sempre um impacto positivo, mas este depende da variável que estamos a acrescentar. Acrescentar a variável DAConsumo ou SAConsumo ao conjunto de variáveis de entrada tem um impacto muito mais forte que acrescentar a variável Temperatura, por exemplo, o que vai de acordo com a dimensão dos índices de IM destas variáveis.

Existem também 2 pontos, que se encontram marcados na figura 12, que apresentam resultados particularmente surpreendentes, considerando não só o somatório dos índices das variáveis envolvidas, como também o número de variáveis. Trata-se dos resultados das redes neuronais obtidas utilizando {Hora, DS, DAConsumo} e {Hora, DS, DAConsumo, Temperatura}. Estas duas redes obtiveram resultados muito bons, sem ser necessário sobrecarregar no número de entradas. Em ambos os conjuntos encontra-se 3 variáveis em comum, {Hora, DS, DAConsumo}. Haverá algum relacionamento entre estas três variáveis que, quando utilizadas em conjunto, levem diretamente a uma boa previsão, sem sobrecarregar o conjunto de entrada com variáveis?

5.3 - DAConsumo vs SAConsumo

Durante todas as análises que se realizaram aos índices de IM, a variável SAConsumo sempre apresentou sinais de estar mais correlacionada do que a variável DAConsumo, relativamente à nossa variável de saída, o Consumo. Apesar da variável DAConsumo ser referente a um período mais recente no tempo, apenas 2 dias de distância, as diferenças no diagrama de carga entre fins-de-semana e dias da semana são uma das razões porque a variável DAConsumo apresente, pelo que temos visto, piores resultados que a variável SAConsumo, que é referente a um período equivalente ao que estamos a prever, não sendo afetada pelas diferenças no diagrama de carga semanais.

No entanto, no último capítulo chegamos a resultados onde a variável DAConsumo ultrapassava, e por uma boa margem, a variável SAConsumo, em termos de performance.

Analisemos as linhas 2 e 3 da tabela 21:

Variáveis	MAPE	MSE	Razão MI
Hora,DS,SAConsumo	3,96%	3580659	71,87%
Hora,DS,DAConsumo	3,32%	2604014	45,05%

Tabela 21 - Comparação entre os índices de IM e o MSE e MAPE: Hora, DS, SAConsumo vs Hora, DS, DAConsumo

Apesar do somatório dos índices de IM ser maior, o conjunto de variáveis {Hora, DS, SAConsumo} acaba por apresentar piores resultados.

No capítulo 5.1 foram analisadas as redes neuronais criadas a partir de cada uma das variáveis escolhidas, onde a variável SAConsumo, com o maior índice de IM de todas as variáveis, apresentou o menor erro. Como será possível que agora seja ultrapassada pela variável DAConsumo.

Analisemos os resultados obtidos pela combinação das várias variáveis:

Variáveis	MAPE	MSE	Razão MI
DS,DAConsumo	5,54%	6682954	22,59%
DS,SAConsumo	4,27%	4435859	49,41%
Hora,DAConsumo	8,39%	13459622	37,53%
Hora,SAConsumo	4,45%	4995423	64,35%
Irradiancia, DAConsumo	9,60%	17214764	19,27%
Irradiancia, SAConsumo	4,46%	5094493	46,09%
Pluviosidade, DAConsumo	10,90%	20880993	15,68%
Pluviosidade, SAConsumo	4,46%	5139584	42,50%
Temp, DAConsumo	9,94%	18088752	22,15%
Temp, SAConsumo	4,28%	4720678	48,97%
DAConsumo, SAConsumo	4,32%	4757306	56,96%

Tabela 22 - Comparação entre os índices de IM e o MSE: SAConsumo vs DAConsumo

Analisando a tabela 23, para todas as comparações apresentadas, a variável SAConsumo apresenta melhores resultados, o que é uma continuação do que sucedeu quando se utilizou cada variável sozinha como conjunto de entrada.

No entanto existem alguns valores que vão contra o que já vimos até aqui, ou seja, que um aumento do somatório dos índices de IM está ligado a um menor erro, ou seja, melhores resultados.

Na última linha da tabela, a utilização de ambas as variáveis DAConsumo e SAConsumo leva a melhores resultados do que a combinação {Hora, SAConsumo}, apesar de possuir um somatório menos dos índices de IM. Não se trata apenas deste caso, pois a combinação de

variáveis {DS, DAConsumo} apresenta melhores resultados que o conjunto de entrada constituído pelas variáveis {Hora, DAConsumo}. Relembrando que o índice que se encontra a ser utilizado para DS é para uma dimensão diferente, menor que os restantes, o facto de mesmo assim pertencer a um conjunto de variáveis de entrada com melhores resultados indica uma sinergia forte com DAConsumo.

Refletindo um pouco mais sobre estes dois casos. Quando são utilizadas ambas as variáveis referentes ao Consumo numa altura anterior, existe muita informação disponível sobre a evolução típica do consumo ao longo do tempo. Assim, visto que os dados já se encontram ordenados cronologicamente, ter informações sobre o consumo em 2 alturas diferentes é muito mais importante que ter informações sobre a Hora, e o consumo num momento passado.

No segundo caso mencionado, as variáveis DS e DAConsumo parecem complementar-se bem e serem capazes de atingir bons resultados numa previsão. Analisemos a relação entre ambas. A variável DAConsumo apresenta os valores do consumo registados, para a mesma hora, 2 dias antes, enquanto que a variável DS nos diz em que dia da semana nos situamos. Juntando estas duas informações, estamos a fornecer à rede a informação necessária de modo a que esta seja capaz de pesar bem os valores, pois fornecemos a possibilidade de obter uma ligação entre o Dia da Semana em que se refere a previsão, e o quanto o valor passado pode estar, ou não, relacionado com o valor que se pretende prever. Por outras palavras, estamos a possibilitar a rede neuronal de conseguir distinguir entre um diagrama de cargas diário verificado a um dia da semana, com um diagrama de cargas registado num fim-de-semana, ou seja, estas duas variáveis contêm uma grande sinergia entre si.

No entanto, ainda não se conseguiu explicar o resultado visto na tabela 22. Quando agrupadas a variável DAConsumo ou com DS, ou com a Hora, o conjunto de variáveis de entrada resultante resulta numa rede neuronal que produz sempre piores resultados do que quando é utilizada a variável SAConsumo em substituição.

Assim sendo, quando as 3 variáveis se juntam, os resultados se invertem. De modo a verificar isto, decidiu-se acrescentar mais uma variável a cada conjunto de entrada, e verificar se esta conclusão se mantém:

Variáveis	MAPE	MSE	Razão MI
Hora, DS, SAConsumo, Temperatura	3,60%	2942514	78,96%
Hora, DS, DAConsumo, Temperatura	3,47%	2589453	52,14%

Tabela 23 - Comparação entre os índices de IM e o MSE e MAPE: SAConsumo vs DAConsumo c/ Temperatura

Não há razões que indiquem que a temperatura influencie mais um conjunto de variáveis

do que o outro, e analisando os resultados comparativamente com a tabela 22, verifica-se isso mesmo.

Mesmo com a adição de uma nova variável, a mesma situação surgiu: DAConsumo faz parte do conjunto de variáveis de entrada que possui os melhores resultados.

Como pudemos ver antes, as variáveis DAConsumo e DS possuem uma boa sinergia, e quando a estas 2 se acrescenta a variável Hora, a rede é capaz de otimizar o relacionamento entre as variáveis, para produzir ainda melhores resultados.

Esta conclusão é muito importante, e é algo que o conceito de informação mútua não nos apresenta (também devido ao facto de apenas trabalharmos com relações entre apenas 2 variáveis, e não 3 ou mais). Assim, apesar de haver indícios que a aplicação do conceito de Informação Mútua esteja ligado aos resultados das redes neuronais, a nossa ferramenta de teste, a relação entre as variáveis será sempre algo que terá que se ter em conta.

5.4 - Teste das variáveis Manipuladas

Neste capítulo será determinado o impacto das transformações que se efetuaram sobre as variáveis Hora e DS, permitindo-nos ter uma ideia se a quantidade de informação contida nestas novas variáveis permite-nos considerá-las como ainda sendo a mesma variável, para propósitos de cálculos de índices de IM, ou se se trata de uma variável completamente nova, e uma possível evolução relativamente às variáveis antigas.

5.4.1 - Teste da variável Hora_168

Seguindo a mesma metodologia utilizada anteriormente, as variáveis serão comparadas inicialmente a nível individual, e só depois como parte de um conjunto de variáveis de entrada. Como a variável Hora_168 foi obtida a partir das variáveis Hora e DS, o primeiro passo passa por comparar os resultados à saída das redes neuronais destas 3 variáveis:

Variável	MAPE	MSE
Hora_168	7,74%	11599583
Hora	9,24%	17140303
DS	13,39%	29462113

Tabela 24 - Valores do MAPE e MSE das variáveis Hora, DS e Hora_168

Como seria expectável, a nova variável apresenta melhores resultados, quando comparada com as variáveis originais. Isto deve-se ao facto da variável conter informações de ambas as variáveis Hora e DS, o que leva a rede neuronal a ser capaz de chegar a um conjunto de saída mais próximo do real.

De modo a verificar o seu impacto num ambiente de previsão mais real, compararam-se os resultados de redes contendo de um lado a variável Hora_168, e do outro, as variáveis Hora e DS:

Variáveis	MAPE	MSE
Hora, DS	6,06%	7748242
Hora,DS,SAConsumo	3,96%	3580659
Hora,DS,DAConsumo	3,32%	2604014
Hora,DS,Temperatura	4,33%	4025539
Hora,DS,DAConsumo,SAConsumo	2,85%	1909901
Hora,DS,Temperatura,Pluviosidade	4,29%	3942885
Hora, DS, DAConsumo,SAConsumo,Temperatura	2,63%	1624445
Hora, DS, DAConsumo,SAConsumo,Temperatura,Irradiância	2,54%	1515141

Tabela 25 - Valores do MAPE e MSE p/ conjuntos de entrada c/ Hora e DS

Variáveis	MAPE	MSE
Hora_168,SAConsumo	4,30%	4524955
Hora_168,DAConsumo	4,04%	3670748
Hora_168,Temperatura	8,64%	14324390
Hora_168,DAConsumo,SAConsumo	3,04%	2144676
Hora_168,Temperatura, Pluviosidade	8,41%	14318844
Hora_168, DAConsumo, SAConsumo, Temperatura	3,12%	2313479
Hora_168,DAConsumo,SAConsumo,Temp,Irradiância	2,92%	1955657

Tabela 26 - Valores do MAPE e MSE p/ conjuntos de entrada c/ Hora_168

Como regra geral na previsão, o aumento do número de variáveis leva a uma melhoria dos resultados. Isto verifica-se aqui, com os conjuntos de entrada contendo as variáveis Hora e DS a apresentarem melhores resultados que a nova variável. No entanto, na maioria dos casos, a diferença não é muito significativa.

Mais importante ainda, é a comparação direta entre os resultados utilizando a variável Hora_168 e a combinação Hora e DS, sendo que a última apresenta um erro ligeiramente menor.

Algo que é necessário realçar é o facto da nova variável manter a propriedade vista no capítulo 5.3, ou seja, a sua correlação com as variáveis DAConsumo e SAConsumo. Como a variável Hora_168 é construída a partir de ambas as variáveis Hora e DS, pode-se considerar que, num conjunto de entrada constituído por {Hora_168, DAConsumo}, se encontram as 3 presentes e, como se pode ver na tabela acima, é capaz de alcançar melhores resultados que um conjunto de entrada homólogo, mas com SAConsumo.

Face a tudo isto, não é possível considerar a Hora_168 como alternativa à variável Hora, devido à quantidade de informação de outras variáveis que está inserida nela. Pode no entanto ser considerada como substituição do par {Hora, DS}, se o previsor estiver disposto a sacrificar um aumento de erro na previsão, pela utilização de menos uma variável. Não sendo esta uma situação muito provável, a utilização de DS e Hora em conjunto permite atingir melhores resultados do que esta nova variável.

5.4.2 - Teste da variável DS_8h

A variável DS_8h foi construída de forma a podermos contornar a limitação da variável DS, e podermos calcular o índice de Informação Mútua para uma dimensão comum a todas as variáveis, neste caso, para uma dimensão de 24.

A variável Hora foi apenas utilizada como referência, de modo a poder fragmentar cada dia em 3 períodos iguais, portanto não se antevê que esta tenha passado muita informação para a nova variável.

Comparando os resultados obtidos pelas redes neuronais, quando utilizadas apenas as 2 variáveis:

Variável	MAPE	MSE
DS_8h	7,54%	10965177
DS	13,39%	29462113

Tabela 27 - Valores do MAPE e MSE das variáveis DS e DS_8h

A nível individual a nova variável ultrapassa largamente a performance da variável DS, sendo questionável afinal a transferência de informação que a variável Hora terá realizado, aquando a construção desta nova variável.

Para tal, procedeu-se a alguns testes, considerando agora mais do que uma variável no conjunto de entrada das redes neuronais:

Variáveis	MAPE	MSE
Hora, DS	6,06%	7748242
DS_8h	7,54%	10965177

Tabela 28 - Valores do MAPE e MSE das variáveis DS+Hora e DS_8h

Começando a nossa análise por partes. Como se pode verificar na tabela 29, apesar da variável DS_8h ter sido construída utilizando informações de ambas as variáveis Hora e DS, a utilização das mesmas continua a produzir melhores resultados que o simples uso da variável DS_8h. Isto significa que a variável Hora ainda contém muita informação, que ajuda a

encontrar melhores resultados, quando combinada com DS, comparativamente ao uso de apenas DS_8h.

Variáveis	MAPE	MSE
DS,DAConsumo	5,54%	6682954
DS,SAConsumo	4,27%	4435859

Tabela 29 - Valores do MAPE e MSE c/ DS

Variáveis	MAPE	MSE
DS_8h, DAConsumo	4,31%	3887384
DS_8h, SAConsumo	4,25%	4344541

Tabela 30 - Valores do MAPE e MSE c/ DS_8h

Analisando os resultados obtidos através do uso de DS e DS_8h com outras variáveis, tiramos conclusões repartidas. Enquanto que temos uma melhoria de resultados quando as variáveis DS_8h e DAConsumo são utilizadas, em comparação à utilização de DS e DAConsumo, para o caso do SAConsumo, isso não se verifica, ou melhor, não está tão presente: a diferença entre os resultados não é muito significativa.

Existe alguma informação sobre a hora na variável DS_8h, e como já foi possível observar, o trio de variáveis DS, Hora e DAConsumo apresentam normalmente bons resultados.

De modo a tirar as dúvidas, testou-se a nova variável com o mesmo conjunto de variáveis de entrada, mas com a adição da variável Hora:

Variáveis	MAPE	MSE
Hora,DS,SAConsumo	3,96%	3580659
Hora,DS,DAConsumo	3,32%	2604014

Tabela 31 - Valores do MAPE e MSE c/ DS e Hora

Variáveis	MAPE	MSE
Hora, DS_8h, SAConsumo	3,95%	3626936
Hora, DS_8h, DAConsumo	3,31%	2556951

Tabela 32 - Valores do MAPE e MSE c/ DS_8h e Hora

Apesar de alguma diferença no MSE, os resultados são muito próximos. Isto leva-nos ao facto que, juntamente com a variável Hora, as variáveis DS e DS_8h têm o mesmo comportamento com as variáveis DAConsumo e SAConsumo. De modo a testar isto, considerou-se um novo caso, em que são consideradas as 4 variáveis, juntamente com a variável Temperatura:

Variáveis	MAPE	MSE
Hora, DS, DAConsumo,SAConsumo,Temperatura	2,63%	1624445
DS_8h, DAConsumo, SAConsumo, Temperatura	2,96%	1988953
Hora, DS_8h, DAConsumo, SAConsumo, Temperatura	2,59%	1558290

Tabela 33 - Valores do MAPE e MSE para comparação de DS e DS_8h

Novamente os resultados são muito próximos, quando é considerada a Hora, em adição às restantes variáveis.

Com isto podemos concluir que, num frente-a-frente, a variável DS_8h é capaz de fornecer melhores resultados que a variável DS, não devendo ser considerada a mesma variável para usos de cálculo de índices de IM, por exemplo. No entanto, se a variável Hora e pelo menos 1 das variáveis referentes ao consumo num período anterior for utilizada, então as variáveis DS e DS_8h têm o mesmo comportamento, sendo praticamente indiferente qual das duas deverá ser utilizada (a variável DS_8h continua a apresentar resultados melhores, ainda só que ligeiramente).

5.5 - Teste das variáveis Meteorológicas

No capítulo 4.5 os índices de Informação Mútua obtidos para as variáveis meteorológicas (Vel. Vento, Irradiância, Pluviosidade e Temperatura) eram muito idênticos aos obtidos para as mesmas variáveis, mas 2 dias antes (DAVel. Vento, DA Irradiância, DA Pluviosidade e DATemperatura). Resta agora saber se as redes neuronais também vão apresentar resultados idênticos à saída.

A metodologia usada foi a mesma utilizada anteriormente, analisando primeiro os resultados obtidos por cada variável isolada, e depois conjuntos de variáveis de entrada com mais do que uma variável.

Variáveis	MAPE	MSE
Irradiância	14,39%	32207475
Temperatura	15,47%	36192405
Vel. Vento	15,44%	35898305
Pluviosidade	14,32%	32580221

Tabela 34 - Valores do MAPE e MSE das variáveis meteorológicas

Variáveis	MAPE	MSE
DA Irradiância	14,22%	32503955
DATemperatura	14,42%	32096694
DAVel. Vento	15,38%	35915054
DA Pluviosidade	15,47%	36273910

Tabela 35 - Valores do MAPE e MSE das variáveis meteorológicas 2 dias antes

Tal como os valores dos índices, as variáveis apresentam valores muito semelhantes à saída das redes neuronais, no que toca aos erros. No entanto, as variáveis DA Pluviosidade e Vel. Vento não apresentam a sua dominância perante a sua “variável equivalente”, ao apresentar erros superiores.

Devido à natureza destas variáveis, foi decidido utilizar a variável Hora, pois é uma variável que apresenta uma ligação com todas estas variáveis de uma forma ou outra.

Variáveis	MAPE	MSE
Hora,Irradiância	9,23%	16605269
Hora,Temperatura	8,44%	13386936
Hora,Vel. Vento	9,16%	16746919
Hora,Pluviosidade	9,11%	16887226

Tabela 36 - Valores do MAPE e MSE das variáveis meteorológicas c/ Hora

Variáveis	MAPE	MSE
Hora,DA Irradiância	9,12%	16316394
Hora,DATemperatura	8,47%	13528450
Hora,DAVel. Vento	9,11%	16633598
Hora,DA Pluviosidade	9,09%	16518617

Tabela 37 - Valores do MAPE e MSE das variáveis meteorológicas 2 dias antes c/ Hora

Com a introdução da variável Hora, as diferenças dos erros obtidos à saída das redes neuronais é ainda menor. De realçar que, com esta adição, todas as variáveis DA apresentam melhores resultados que as variáveis correspondentes ao tempo real.

A conclusão importante a retirar é a semelhança de resultados na utilização de umas ou outras variáveis, confirmando os valores obtidos para os índices destas variáveis.

Capítulo 6

Conclusão

Após a realização deste trabalho torna-se mais fácil entender a importância do processo de seleção de variáveis, e como este pode afetar a previsão, ainda antes de esta ser efetuada.

De modo a conseguir chegar a um compromisso entre performance e número de variáveis de entrada, é fulcral compreender a importância/peso que uma dada variável tem no conjunto de saída, assim como de que forma é que as variáveis se relacionam entre si. O primeiro ponto é mais fácil, existindo várias ferramentas, algumas delas referidas ao longo deste trabalho, incluindo o conceito aplicado de Informação Mútua, que ajudam a determinar uma medida do peso individual de cada uma das variáveis. A forma de como as variáveis se relacionam pode já ser algo completamente diferente e de análise bastante mais complexa. É relativamente fácil perceber a sinergia entre as variáveis Hora e Dia da Semana, ou entre DAConsumo e SAConsumo pois fornece, no primeiro caso, uma referência para o tempo e, no segundo, informações sobre o comportamento da saída no passado. Contudo, ao trabalhar com conjuntos de variáveis de entrada, podem existir relações entre elas que afetam largamente o resultado à saída da previsão. O caso analisado das variáveis DAConsumo e SAConsumo é um exemplo disso, em que a variável SAConsumo apresenta sempre melhores resultados que DAConsumo, exceto se nos conjuntos das variáveis de entrada estiver presente ambas Hora e Dia da Semana, sendo que a partir daí, DAConsumo apresentará constantemente melhores resultados.

O conceito de Informação Mútua, apesar dos cuidados necessários exigidos no cálculo dos índices, permitiu obter indicadores consistentes com os resultados dos testes de previsão efetuados.

Não existem métodos perfeitos para seleção de variáveis, e as suas limitações com a utilização de variáveis discretas é prova disso para este caso. No entanto, recorrendo a alternativas, ou através da transformação de variáveis, ou aproximação de resultados, é

possível ultrapassar este obstáculo. No entanto, as vantagens contrastam significativamente com os obstáculos, sendo que este método permite obter uma ideia do impacto que certas variáveis irão ter no decorrer da previsão, permitindo ao utilizador antever uma relação entre o número de variáveis a utilizar e o aumento da performance da previsão, com muito mais facilidade. Permite também a diminuição do número de variáveis candidatas, excluindo aquelas com um índice muito baixo, diminuindo assim o número de combinações possíveis a testar, o que é um grande ganho de tempo e esforço.

No geral, a aplicação deste conceito passa pelos seguintes passos:

- Tratamento da informação - Igualar o tamanho da amostra de todas as variáveis, pois quantidades de dados diferentes apresentam quantidades de informação diferentes, o que irá afetar os valores dos índices;
- Cálculo da Matriz das Probabilidades entre a variável de teste e a variável de saída;
- Cálculo dos índices de Informação Mútua;
- Cálculo das importâncias relativas.

Analisando o panorama geral, o conceito de IM tem potencial para ser uma ferramenta muito poderosa no processo de seleção de variáveis, não só face aos resultados capaz de apresentar, mas devido à baixa capacidade de computação exigida, o que torna o método bastante acessível.

Vale realmente utilizar este método? Esta será talvez a pergunta mais pertinente. Se o conjunto de variáveis a avaliar for composto maioritariamente de variáveis discretas, então este poderá não ser o melhor método a ser utilizado, salvo seja possível construir uma matriz das probabilidades de dimensão comum a todas as variáveis. Se o conjunto de variáveis for composto maioritariamente de variáveis contínuas, então este é um excelente método de avaliação do potencial de cada variável, e os resultados da previsão deverão coincidir com os valores dos índices obtidos para as variáveis.

Referências

- [1] Variáveis Endógenas vs. Exógenas -
http://www.fep.up.pt/docentes/tina/Microeconomia1/conceitos_recordear.pdf
- [2] Conceitos teóricos de Entropia e Informação Mútua -
<http://homes.cs.washington.edu/~anuprao/pubs/CSE533Autumn2010/lecture3.pdf>
- [3] Definição e propriedades do conceito de Informação Mútua -
http://www.scholarpedia.org/article/Mutual_information
- [4] Entropia e a 2ª Lei da Termodinâmica -
<http://sistemas.eel.usp.br/docentes/arquivos/2166002/LOB1019/Fisica2.Cap20Entropia.TERCEIRAVALIACA0.pdf>
- [5] Previsão - <http://en.wikipedia.org/wiki/Forecasting>
- [6] Entropia e a 2ª Lei da Termodinâmica - <http://www.e-escola.pt/topico.asp?id=573>
- [7] Entropia (Teoria da Informação) -
http://en.wikipedia.org/wiki/Entropy_%28information_theory%29
- [8] Teoria da Informação -
http://en.wikipedia.org/wiki/Information_theory_and_measure_theory#Multivariate_mutual_information
- [9] Construção da Matriz das Probabilidades -
<http://scienceonthesquares.blogspot.pt/2012/11/introduction-to-mutual-information.html>
- [10] Conceitos de Entropia e Informação Mútua -
<https://web.cse.msu.edu/~cse842/Papers/CoverThomas-Ch2.pdf>

- [11] Entropia e a Termodinâmica - <http://sistemas.eel.usp.br/docentes/arquivos/2166002/LOB1019/Fisica2.Cap20Entropia.TERCEIRAVALIACA0.pdf>
- [12] Conceitos da Informação - <http://courses.ischool.berkeley.edu/i218/s09/slides/Cofi09-Shannon%28PD%29.pdf>
- [13] Construção da Matriz das Probabilidades - http://www.maxwell.lambda.ele.puc-rio.br/12170/12170_5.PDF
- [14] Previsão de carga a curto prazo através de um rede neuronal com treino baseado na entropia - <http://www.sba.org.br/rsv/SBAI/SBAI2011/82672.pdf>
- [15] Claude Shannon, “A Mathematical Theory of Communication”, Julho - Outubro 1948 - <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>
- [16] Seleção de Variáveis - <http://www.portalaction.com.br/887-sele%C3%A7%C3%A3o-de-vari%C3%A1veis>
- [17] Seleção de Variáveis - <http://www.biostat.jhsph.edu/~iruczins/teaching/jf/ch10.pdf>
- [18] Informação Mútua c/ Múltiplas Variáveis - http://en.wikipedia.org/wiki/Multivariate_mutual_information
- [19] Revisão sobre Informação Mútua c/ Múltiplas Variáveis - <https://www3.nd.edu/~jnl/ee80653/Fall2005/tutorials/sunil.pdf>
- [20] Introdução às Redes Neurais - http://www.professores.uff.br/jmarcos/index.php?option=com_content&view=article&id=11&Itemid=26
- [21] Redes Neurais Artificiais - <http://www.icmc.usp.br/pessoas/andre/research/neural/>
- [22] Introdução às Redes Neurais Artificiais - http://www.inf.ufrgs.br/~danielnm/docs/intro_rna.pdf

[23] Redes Neurais [Conteúdos da U.C. de Técnicas para Previsão da FEUP] -

https://sigarra.up.pt/feup/pt/conteudos_service.conteudos_cont?pct_id=136809&pv_cod=12LaxaYm6laz

[24] Redes Neurais [Conteúdos da U.C. de Técnicas para Previsão da FEUP] -

https://sigarra.up.pt/feup/pt/conteudos_service.conteudos_cont?pct_id=202778&pv_cod=12r8NI5EZn2y

Anexos

Nesta secção são apresentados os scripts de MatLab utilizados ao longo deste trabalho.

Anexo A

Script de MatLab para a construção da Matriz das Probabilidades e conseguinte índice de Informação Mútua. O código poderia ser mais compacto, mas deste modo facilita a introdução de dados e deteção de erros.

```
input=[pluv consumo]; %variaveis de entrada
len=length(consumo);
count=0;

min_x=min(pluv); %valor minimo da entrada
max_x=max(pluv); %valor maximo da entrada
nr_x=24; %nr de linhas a colocar na tabela (entradas)
passo_x=(max_x-min_x)/nr_x; %para variáveis contínuas
%passo_x=1; %para variáveis discretas

min_y=min(consumo); %valor minimo da saida (consumo)
max_y=max(consumo); %valor maximo da saida (consumo)
nr_y=24; %nr de columnas a colocar na tabela (consumo)
passo_y=(max_y-min_y)/nr_y;

min_iter_x=min_x;
max_iter_x=min_iter_x+passo_x;

min_iter_y=min_y;
max_iter_y=min_iter_y+passo_y;

m_prob=[]; %matriz das probabilidades

for (j=1:nr_y)
for(i=1:nr_x)
for(k=1:len)
    if(input(k,1)>=min_iter_x) && (input(k,1)<max_iter_x) &&
(input(k,2)>=min_iter_y) && (input(k,2)<max_iter_y)
        count=count+1;
    end
end
valor=count/len;
m_prob(i,j)=valor;
count=0;

min_iter_x=max_iter_x;
max_iter_x=min_iter_x+passo_x;
```

```

end

min_iter_x=min_x;
max_iter_x=min_iter_x+passo_x;

min_iter_y=max_iter_y;
max_iter_y=min_iter_y+passo_y;
end

%calcular MI

I=0;
som_x=sum(m_prob,2);
som_y=sum(m_prob,1);

for(j=1:nr_y)
for(i=1:nr_x)

    quo=som_x(i)*som_y(j);

    if(quo==0 | m_prob(i,j)==0)
        valor=0;
    else
        valor=log10(m_prob(i,j)/quo);
    end

    I=I+(m_prob(i,j)*valor);
end
end

```


Anexo B

Script de MatLab utilizado na criação da variável Hora_168, a partir das variáveis Hora e DS:

```
l1=length(ds);  
hora_168=[];  
  
for(i=1:l1)  
    if(ds(i)==1)  
        hora_168(i)=hora(i);  
    elseif(ds(i)==2)  
        hora_168(i)=24+hora(i);  
    elseif(ds(i)==3)  
        hora_168(i)=48+hora(i);  
    elseif(ds(i)==4)  
        hora_168(i)=72+hora(i);  
    elseif(ds(i)==5)  
        hora_168(i)=96+hora(i);  
    elseif(ds(i)==6)  
        hora_168(i)=120+hora(i);  
    elseif(ds(i)==7)  
        hora_168(i)=144+hora(i);  
    elseif(ds(i)==8)  
        hora_168(i)=hora(i);  
    end  
end
```

Anexo C

Script de MatLab utilizado na criação da variável DS_8h, através da utilização das variáveis

Hora e DS:

```
m_in=[ds hora];
ll=length(ds);
ds_24=[];
for(i=1:ll)
    if(ds(i)==1)
        if(m_in(i,2)>=0 & m_in(i,2)<=7)
            ds_24(i)=0;
        elseif (m_in(i,2)>=8 & m_in(i,2)<=15)
            ds_24(i)=1;
        elseif(m_in(i,2)>=16 & m_in(i,2)<=23)
            ds_24(i)=2;
        end

    elseif(ds(i)==2)
        if(m_in(i,2)>=0 & m_in(i,2)<=7)
            ds_24(i)=3;
        elseif (m_in(i,2)>=8 & m_in(i,2)<=15)
            ds_24(i)=4;
        elseif(m_in(i,2)>=16 & m_in(i,2)<=23)
            ds_24(i)=5;
        end

    elseif(ds(i)==3)
        if(m_in(i,2)>=0 & m_in(i,2)<=7)
            ds_24(i)=6;
        elseif (m_in(i,2)>=8 & m_in(i,2)<=15)
            ds_24(i)=7;
        elseif(m_in(i,2)>=16 & m_in(i,2)<=23)
            ds_24(i)=8;
        end

    elseif(ds(i)==4)
        if(m_in(i,2)>=0 & m_in(i,2)<=7)
            ds_24(i)=9;
        elseif (m_in(i,2)>=8 & m_in(i,2)<=15)
            ds_24(i)=10;
        elseif(m_in(i,2)>=16 & m_in(i,2)<=23)
            ds_24(i)=11;
        end

    elseif(ds(i)==5)
        if(m_in(i,2)>=0 & m_in(i,2)<=7)
            ds_24(i)=12;
        elseif (m_in(i,2)>=8 & m_in(i,2)<=15)
            ds_24(i)=13;
        elseif(m_in(i,2)>=16 & m_in(i,2)<=23)
            ds_24(i)=14;
        end

    elseif(ds(i)==6)
        if(m_in(i,2)>=0 & m_in(i,2)<=7)
            ds_24(i)=15;
        elseif (m_in(i,2)>=8 & m_in(i,2)<=15)
```

```
ds_24(i)=16;
elseif(m_in(i,2)>=16 & m_in(i,2)<=23)
ds_24(i)=17;
end

elseif(ds(i)==7)
    if(m_in(i,2)>=0 & m_in(i,2)<=7)
        ds_24(i)=18;
    elseif (m_in(i,2)>=8 & m_in(i,2)<=15)
        ds_24(i)=19;
    elseif(m_in(i,2)>=16 & m_in(i,2)<=23)
        ds_24(i)=20;
    end

elseif(ds(i)==8)
    if(m_in(i,2)>=0 & m_in(i,2)<=7)
        ds_24(i)=21;
    elseif (m_in(i,2)>=8 & m_in(i,2)<=15)
        ds_24(i)=22;
    elseif(m_in(i,2)>=16 & m_in(i,2)<=23)
        ds_24(i)=23;
    end
end
end
ds_24=ds_24';
```

